

America's choice

The values of scientific enquiry, rather than any particular policy positions on science, suggest a preference for one US presidential candidate over the other.

The election of a US president almost always seems like a crossroads, but the choice to be made on 4 November feels unusual, and daunting, in its national and global significance.

Science and the research enterprise offer powerful tools for addressing key challenges that face America and the world, and it is heartening that both John McCain and Barack Obama have had thoughtful things to say about them. Obama has been more forthcoming in his discussion of research goals (see *Nature* 455, 446–449; 2008), but both have engaged with the issues. McCain deserves particular credit for taking a stance on carbon emissions that is at odds with that of a significant proportion of his party.

There is no open-and-shut case for preferring one man or the other on the basis of their views on these matters. This is as it should be: for science to be a narrow sectional interest bundled up in a single party would be a terrible thing. Both sides recognize science's inspirational value and ability to help achieve national and global goals. That is common ground to be prized, and a scientific journal's discussion of these matters might be expected to stop right there.

But science is bound by, and committed to, a set of normative values — values that have application to political questions. Placing a disinterested view of the world as it is ahead of our views of how it should be; recognizing that ideas should be tested in as systematic a way as possible; appreciating that there are experts whose views and criticisms need to be taken seriously: these are all attributes of good science that can be usefully applied when making decisions about the world of which science is but a part. Writ larger, the core values of science are those of open debate within a free society that have come down to us from the Enlightenment in many forms, not the least of which is the constitution of the United States.

On a range of topics, science included, Obama has surrounded himself with a wider and more able cadre of advisers than McCain. This is not a panacea. Some of the policies Obama supports — continued subsidies for corn ethanol, for example — seem misguided.

The advice of experts is all the more valuable when it is diverse: 'groupthink' is a problem in any job. Obama seems to understand this. He tends to seek a range of opinions and analyses to ensure that his own opinion, when reached, has been well considered and exposed to alternatives. He also exhibits pragmatism — for example in his proposals for health-care reform — that suggests a keen sense for the tests reality can bring to bear on policy.

Some will find strengths in McCain that they value more highly than the commitment to reasoned assessment that appeals in Obama. But all the signs are that the former seeks a narrower range of advice. Equally worrying is that he fails to educate himself on crucial matters; the attitude he has taken to economic policy over many years is at issue here. Either as a result of poor advice, or of advice inadequately considered, he frequently makes decisions that seem capricious or erratic. The most notable of these is his ill-considered choice of Sarah Palin, the Republican governor of Alaska, as running mate. Palin lacks the experience, and any outward sign of the capacity, to face the rigours of the presidency.

The Oval Office is not a debating chamber, nor is it a faculty club. As anyone in academia will know, a thoughtful and professorial air is not in itself a recommendation for executive power. But a commitment to seeking good advice and taking seriously the findings of disinterested enquiry seems an attractive attribute for a chief executive. It certainly matters more than any specific pledge to fund some particular agency or initiative at a certain level — pledges of a sort now largely rendered moot by the unpredictable flux of the economy.

This journal does not have a vote, and does not claim any particular standing from which to instruct those who do. But if it did, it would cast its vote for Barack Obama. ■

"The core values of science are those of open debate within a free society that have come down to us from the Enlightenment."

Growing stronger

Science in developing countries can withstand the current economic climate.

With the news dominated by failing banks and falling stock markets, it is easy to forget that some indices are moving upwards — such as those that record the strength of science, technology and innovation in developing nations. Despite the turmoil, this is likely to continue.

Snapshots of this trend are captured this week in *A World of Science in the Developing World*, a publication by TWAS, the academy of

sciences for the developing world, which celebrates its 25th anniversary next month (copies are distributed with this issue to subscribers, and are freely available at www.nature.com/twas). In the past 25 years, it is not just the larger countries such as Brazil, China, India and Mexico where the volume and the quality of scientific research have been transformed. Smaller countries such as Chile, Malaysia, Rwanda and Vietnam all regard investing in new knowledge, technology and higher education as national priorities.

The reasons are not hard to spot: most countries of the developed world invest 2% or more of their national incomes on research and development (R&D). At the same time, they enjoy some of the highest standards of living in the world. The link between R&D spending and national wealth is subject to much debate among

social scientists (see, for example, D. Edgerton *Nature* **455**, 1030–1031; 2008). But many funders and science policy-makers see the relationship in more straightforward terms: more science eventually provides valuable social returns, including greater wealth, education and a skilled workforce.

Even the oil-producing states of the Middle East — which have enjoyed good living standards without spending on science and technology — are coming round to this view. The largest-ever university-building programme in the region's history is now under way, financed for the most part by the extra income from high oil prices.

As the price of oil (and other commodities) drops along with confidence in the markets, the strategy of investing more in science is going to be tested to the hilt. The larger developing countries should be able to weather the storm — in part because the state is a relatively larger player in research, compared with more developed countries. If private-sector R&D begins to decline, it should still be possible for the state to maintain investment at — or near — current levels.

Smaller (and poorer) countries will have a harder time. Many are

already borrowing from institutions such as the International Monetary Fund (IMF), and their priority will be to increase spending on welfare programmes. But those that want to stay on the present path have more grounds for hope than was the case in past economic crises.

In the 1980s, for example, countries that borrowed from the IMF and from the World Bank were told to cut back on government spending. Science and higher education were two of the biggest casualties, and in a few countries the effects can still be seen. In 2008, the mood couldn't be more different. In their response to the financial crisis, many developed world governments are talking not of cutting back, but of rebuilding and strengthening public institutions. In addition, there is a consensus that the world's financial architecture needs reforming too. Given such an environment, it will be difficult to justify asking poorer countries to cut back on their knowledge infrastructure when richer countries will be doing the opposite.

There is no doubt that the next few years will be difficult, but for the countries of the developing world, the answer to their present problems is not to undo the gains of the past quarter-century. ■

A risk worth taking

An experiment by the Gates Foundation is food for thought for other research agencies.

Barry Marshall and Robin Warren's unorthodox idea that the bacterium *Helicobacter pylori* was involved in gastritis and peptic ulcers met with ingrained resistance from the guardians of stomach-acid wisdom in the 1980s. Against the odds, the two prevailed, revolutionizing care of the disease and receiving a Nobel prize in 2005. But how many other potentially groundbreaking ideas are dragged down a dark alley and quietly strangled by overly conservative peer review of grant proposals?

Research funding should strive for a balanced portfolio that includes both safe investments and higher-risk work. While the world's financial system has been inflated with wildly excessive risk, research funding has had the opposite problem — exacerbated by ever greater competition for limited funds, it is overly wedded to safe, unadventurous research. This, in effect, ostracizes off-the-wall ideas, which often cross disciplinary boundaries and would have potentially big payoffs should they work. Researchers long ago learned that the last people they should tell about their big ideas are their sources of financial support.

To be fair, there are exceptions to such conservatism. The US National Institutes of Health, for example, has systematically promoted risky research through several initiatives (see <http://nihroadmap.nih.gov/highrisk>), although the jury is out on its scientific impact. Military imperatives for innovation led the US Defense Advanced Research Projects Agency to transgress the conventions of academic business-as-usual, with undisputed success.

In a similar vein, frustrated with repeatedly encountering the same faces from the same countries, and a narrow range of ideas, the Bill & Melinda Gates Foundation decided to cast the net wider in the search for new people and ideas. Last week, it announced the 104 winners of the first round of its five-year, US\$100-million 'Grand Challenges

Explorations' programme. This solicited unconventional ideas for protecting against infectious diseases, limiting drug resistance and exploring latent tuberculosis. Future rounds will include ideas for vaccines for killer diseases and tools to help eradicate malaria.

It is the nature of the call for proposals, and their peer review, that is intriguing. The grant proposal is one that many researchers can only dream of — a two-page explanation of the idea, with no supporting data required. To emphasize that it's the idea that matters, reviewers were blinded to the name, profession and nationality of the applicant.

The reviewers themselves were atypical. Instead of consensus review by experts in the field — as is the norm — the 4,000-odd proposals received were sent to individuals, not just in science but also in engineering, business and beyond — people the foundation considers to be 'champions' with strong track records in high-risk research.

It is too soon to assess the projects' chances of success, but what is clear is that many of the projects and their proponents are very different from those in conventional global-health research. They include work to explore links between natural HIV-resistance and genetic markers for type 2 diabetes, a proposal from an astrophysicist to reduce transmission of malaria with a flashlight that would disorientate mosquitoes, and testing the bacterium *Bdellovibrio*, which predated on many Gram-negative pathogens, as a living antibiotic.

The first-round grants, at \$100,000, are small. Dishing out large sums of money on far-fetched ideas would be foolhardy, given that as many as nine out of ten of these projects are expected to fail. But those that show signs of success will be eligible for further funds of \$1 million, or much more. All told, however, the entire programme amounts to just 10% of the Gates Foundation's new Global Health Discovery wing, which itself accounts for just 20% of the organization's overall global-health research budget.

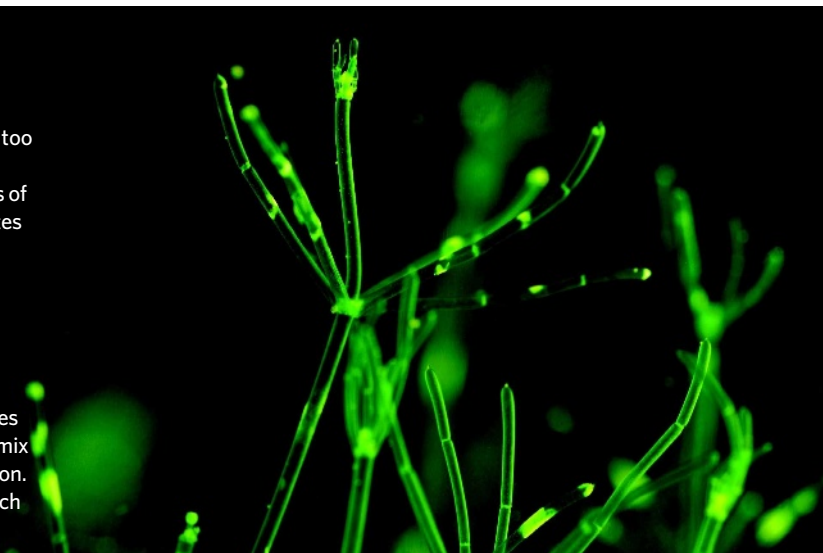
Such levels of investment seem sensibly aligned with levels of risk. Risk-taking brings new faces and ideas to the table at reasonable cost and spurs creativity. Research agencies everywhere need to take a hard look at their funding portfolios to ensure that they are investing sufficiently in high-risk and potentially transformative research. ■

RESEARCH HIGHLIGHTS

Turn of the screw*Phys. Rev. Lett.* **101**, 178102 (2008)

The tubular cells of the aquatic plant *Chara corallina* are several centimetres long and about 1 millimetre wide — too big for diffusion to effectively disperse nutrients within them, but too small for the pumping circulatory systems of large organisms. So the cytoplasm of such a cell circulates by advection. This is driven by the movement of myosin motor proteins along 'tracks' of the filamentous protein actin on the cell's inner walls. But the details of how this works have been unclear.

Jan-Willem van de Meent and his colleagues at the University of Cambridge, UK, now present a model in which advection by the myosin 'conveyor belts' generates two counterflowing, helically entwined vortices. These mix and homogenize solutes much faster than would diffusion. The idea could be copied in microfluidic systems, in which generating rapid mixing is difficult.



J.-W. VAN DE MEENT, M. POLIN, R. E. GOLDSTEIN/UNIV. CAMBRIDGE

ZOOLOGY**Hot and bothered***Proc. Natl Acad. Sci. USA* doi:10.1073/pnas.0806830105 (2008)

Like many spiders, *Erigone atra*, which tends to live among crops, has a choice to make when it decides to move. It can either spin a silk sail to 'balloon' a relatively long way, or spin a rope and rappel a shorter distance.

A team led by Dries Bonte of Ghent University in Belgium has discovered that the temperature that *E. atra* experiences during development influences the means of transport it chooses. The researchers raised 520 spiders in Petri dishes and then placed them on platforms in a wind tunnel. Spiders raised under cool, spring-like conditions tended to balloon, whereas those raised in summer-like conditions preferred to rappel.

MOLECULAR BIOLOGY**Stemming senescence***Cell* **135**, 227–239 (2008)

Researchers have uncovered a biochemical pathway involving tumour suppressor proteins that changes the expression of genes in neural stem cells as the cells age. This casts the ebbing regenerative capacity of ageing tissues as a trade-off that keeps cancer at bay.

Sean Morrison at the University of Michigan in Ann Arbor and his colleagues worked out that the decrease in Hmga2, a regulator of gene expression, that occurs as neural stem cells age allows levels of the tumour suppressors p16^{Ink4a} and p19^{Arf} to increase. This leads to lower self-renewal rates. The researchers counted the number of neural cells per section of intestine (pictured right)

as one measure of self-renewal.

Hmga2 promotes cell renewal in young mice by maintaining low tumour suppressor levels. In old mice, however, avoiding cancer takes precedence over keeping neural matter youthful.

CHEMICAL BIOLOGY**NO problem***Nature Chem. Biol.* doi:10.1038/nchembio.115 (2008)

Nitric oxide (NO) is involved in a host of mammalian signalling pathways, and several enzymes are charged with its synthesis. The overactivity of iNOS, one of these enzymes, has been linked to multiple sclerosis and Alzheimer's disease. Now researchers have found a way to inhibit it.

Elsa Garcin and Elizabeth Getzoff at the Scripps Research Institute in La Jolla, California, and their colleagues reasoned that the best approach to selectively block the

action of this enzyme would be to study the structures of two drugs that preferentially hinder iNOS over other NO-producing enzymes. They then designed a new iNOS inhibitor that anchors itself to a binding pocket on the enzyme before inducing a cascade of conformational changes that make the inhibitor's grip more secure.

CONDENSED-MATTER PHYSICS**Sound trap***Nature Phys.* doi:10.1038/nphys1101 (2008)

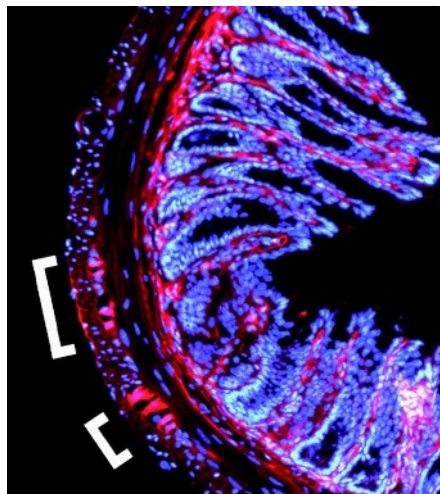
Since the 1950s, physicists have known that a very disordered material can trap electrons. This is because of the electron's wave behaviour. When the degree of disorder is sufficient, the electron waves can become 'localized' in a single spot.

This trapping behaviour is complex, but John Page at the University of Manitoba in Winnipeg, Canada, and his colleagues have created an analogous but simpler system for sound waves. They fused small aluminium ball-bearings together and, using ultrasound, were able to observe sound waves become trapped in the disordered aluminium structures. They believe that their work may improve the understanding of localization, which could have applications in electronics.

ZOOLOGY**Born to reign***Science* **322**, 552 (2008)

Harvester ants develop into queens or sterile workers according to their genes as well as their environment, which had been thought to be the lone determinant of female caste.

Tanja Schwander, currently at Simon



Fraser University in Burnaby, Canada, and Laurent Keller of the University of Lausanne in Switzerland noticed that some combinations of genes are more common in queen harvester ants (*Pogonomyrmex rugosus*) than in workers, and vice versa. So they crossed queens from six colonies with males of the same and different colonies to see what effect this might have.

The duo found that certain cross-colony pairings — and therefore particular gene combinations — reliably skewed the proportion of queens to female workers compared with the average.

NEUROSCIENCE

Twitchy details

Cell **135**, 334–342 (2008); *Neuron* **60**, 285–297 (2008)
The connections between nerves and the muscles that enable animals to move and breathe cannot form without a protein called Agrin that activates an enzyme called MuSK. Two research groups have discovered that Agrin binds to Lrp4, a receptor, and together they switch on MuSK.

Steven Burden and a team from New York University Medical School stimulated normal and Lrp4-deficient developing muscle fibres with Agrin. MuSK was only activated in cells with Lrp4. The team then infected the mutant cells with a virus carrying the Lrp4 gene, which caused these cells to start producing the receptor. Agrin could then turn on MuSK.

Lin Mei, of the Johns Hopkins University School of Medicine in Baltimore, Maryland, and his co-workers linked Lrp4 expression to the clustering of neurotransmitter receptors at the junction of nerve and muscle cells.

MOLECULAR BIOLOGY

On the mend

Science **322**, 597–602 (2008)
Some sorts of DNA damage tend to get repaired right at the edge of a cell's nucleus, Susan Gasser of the Friedrich Miescher Institute for Biomedical Research in Basel, Switzerland, and her colleagues have shown.

They followed the process of DNA repair by inducing a DNA double-strand break and tracking fluorescent proteins tagged near the site of the break. Damaged DNA was shunted to the edge of the nucleus — to a subcomplex of the nuclear pore that facilitates the exchange of RNA and proteins between the cytoplasm and the cell nucleus. The

transport involved repair proteins that are bound to a protein called SUMO, and are also recognized by a pore-associated ubiquitin ligase, a type of enzyme.

ECOLOGY

Cone heads

Proc. R. Soc. B doi:10.1098/rspb.208.0742 (2008)
The affinity of insects for bright lights is well known, but at least one species seems to 'see' the heat of seed cones (pictured below, right image taken with infra-red camera) from afar.

Gerhard Gries of Simon Fraser University in Burnaby, Canada, and his colleagues have identified infra-red (IR) receptors on conifer seed bugs (*Leptoglossus occidentalis*), and shown with the aid of IR sources in the lab and the field that these creatures prefer hot things. The insects displayed a preference for traps that emitted strong rather than weak IR — unless their IR receptors were covered with silica paint.



ASTRONOMY

Planet probing

Astrophys. J. **686**, 1341–1348 (2008)
A tiny bit of light from an extrasolar 'hot Jupiter' has allowed astronomers to speculate about the diversity of the atmospheres of similar planets.

David Charbonneau of the Harvard-Smithsonian Center for Astrophysics in Cambridge, Massachusetts, and his colleagues used the Spitzer Space Telescope to observe a relatively nearby hot Jupiter before and after a 'secondary transit'. These occur as a planet goes behind a star. The number of photons detected before and after the eclipse is compared with the number detected during the eclipse, and the difference is the number that can be attributed to the planet itself.

Charbonneau's team discerned that the hot Jupiter emitted less light in three spectral bands associated with water than in surrounding bands, contrary to findings from another hot Jupiter. This suggests that there are at least two types of hot-Jupiter atmosphere: one that gets cooler with altitude, and another that has a hot layer high up.

S. TAKACS

JOURNAL CLUB

Bartosz Grzybowski
Northwestern University,
Evanston, Illinois

A physical chemist is pleased to learn that 'microscale' swimming isn't that hard after all.

Even if small organisms perfectly mimicked gold medallist Michael Phelps's technique, they wouldn't win a microswimming Olympics. The viscosity of water is so high that these little fellows have had to develop some unusual swimming styles. In 1977, E. M. Purcell formally expressed this idea with his famous 'scallop theorem'. He showed that swimming forwards cannot be achieved at the micrometre-scale with 'time-reversible' motions such as the back-and-forth wiggling of a rigid tail. Instead, tiny organisms must use complex, asymmetrical strokes.

But this is not always the case, according to engineers at the Massachusetts Institute of Technology in Cambridge and the University of California, San Diego. In July, they proved that time-reversible tail-wiggling or wing-flapping can be a viable mode of propulsion through a fluid, provided it is done next to a deformable interface such as a soft membrane (R. Trouilloud *et al. Phys. Rev. Lett.* **101**, 048102; 2008). The reversible motions of the swimmer couple in a nonlinear way to the deformations of the interface, producing additional flows and forces that are sufficient for locomotion.

One of the most exciting extensions of this result might be in creating 'nanosubmarines' — a much-criticized dream of nanotechnologists to have devices navigate blood vessels, finding and fixing damaged organs as they go. The idea has so far seemed implausible because such machines would need elaborate nanopropellers — which are prohibitively difficult to build — to sculpt asymmetrical swimming motions. But what about using a simpler propulsion mechanism and relying on the deformations of blood-vessel walls to move nanosubmarines along? Is there a nanoshipyard out there somewhere to put this idea to the test?

Discuss this paper at <http://blogs.nature.com/nature/journalclub>

NEWS

THE ELECTION LANDSCAPE

Along with the US presidency, also up for grabs at the 4 November US elections are 470 congressional seats, 11 governor seats and a series of ballot initiatives. Here's a look at some of the races important to science.

CALIFORNIA

Proposition 7 would require the state to produce 40% of its electricity from renewable sources — up from the current mandate of 20% — by 2020. Environmental groups, however, say that the proposal allows too many loopholes for industry.

NEBRASKA

Republican Mike Johanns, the former secretary of agriculture, is favoured to win his Senate race against Democrat rancher Scott Kleeb.

MICHIGAN

Proposal 2 would allow state researchers to derive new human embryonic stem-cell lines (see 'Stem-cell law goes to the polls', below).

ILLINOIS

Bill Foster (Democrat), one of three physicists in the House of Representatives, is favoured to win over Republican businessman Jim Oberweis.

MISSOURI

Proposition C would require investor-owned utilities to generate 15% of their power from renewable sources by 2021.

COLORADO

Amendment 48 would change the definition of 'person' to include human fertilized eggs, embryos and fetuses — potentially outlawing birth-control methods that interfere with the implantation of a fertilized egg.

NEW MEXICO

The retirement of long-time Republican senator Pete Domenici opens up a spot representing some of the country's most powerful nuclear-weapons centres — Los Alamos National Laboratory and Sandia National Laboratories.

Stem-cell law goes to the polls



Last year Sean Morrison, a stem-cell scientist at the University of Michigan in Ann Arbor, received an e-mail from a woman wanting to donate extra embryos from her *in vitro* fertilization procedure towards his research into Parkinson's disease. He had

to say no. According to Michigan law, the woman could donate the embryos to another state or throw them away — but not give them to a local researcher to derive new stem-cell lines.

Michigan's scientific reputation could change on 4 November. Voters will not only choose the next US president (see *Nature* 455, 442–453; 2008) but also say yes or no to a state ballot measure on stem-cell research, known as Proposal 2 — one of several initiatives across the country (see map). If passed, the proposal would amend a 1978 state law banning research on live human embryos, which currently prevents Michigan researchers from deriving new human embryonic

stem-cell lines, and which they see as limiting their research.

The initiative would also mean that, instead of having to discard embryos left over from fertility treatment, women could legally donate them directly to state stem-cell research centres. The state's ban on using somatic-cell nuclear transfer to produce a human embryo would still hold, and buying and selling human embryos would also become illegal.



Bill Clinton (left) and Al Taubman back Michigan's Proposal 2.

Supporters of the initiative had first tried to work within the state legislature to change what they see as overly restrictive laws. But the bills never got out of committee, says Chris De Witt, spokesman for CureMichigan, the group that sponsors Proposal 2. Proponents tried to get a similar initiative on a ballot in 2006, but failed to collect the necessary number of signatures. It's not clear whether the political tides have turned in the initiative's favour this year; a

recent poll by a local newspaper and television stations showed a roughly equal split between those who would vote for the measure and those who would not. It has, however, attracted high-profile supporters such as Michigan governor Jennifer Granholm, billionaire Al Taubman and former president Bill Clinton.

Opponents cite the destruction of human embryos as a reason to vote against the measure. Some groups are specifically worried about part D of the state ballot proposal, which declares it would "prohibit state and local laws that prevent, restrict or discourage stem-cell research, future therapies and cures". That, in essence, would allow future stem-cell research in

C. OSORIO/AP



TOXIC TIDE
Urea pollution may have doomed Hitchcock's kamikaze gulls
www.nature.com/news

UNIVERSAL/THE KOBAL COLLECTION

the state to go unregulated, argues state senator Tom George (Republican), who co-chairs the opposition group, Michigan Citizens Against Unrestricted Science and Experimentation.

That's simply not true, counters De Witt. He notes that state institutions that conduct human embryonic stem-cell research have internal ethics review boards to monitor the studies, and that researchers must also follow federal regulations regarding investigations into humans and human tissue. The same constraints would hold, he says, if researchers were allowed to derive their own lines.

The initiative has "major significance" because presidential candidates John McCain and Barack Obama have both hinted that, if elected, they would loosen federal restrictions on embryonic stem-cell research, says Stephen Rapundalo, executive director of MichBio in Ann Arbor, a non-profit organization trying to drive growth of the life-sciences industry in the state. Currently, federal research money can be used for work only on cell lines derived before

"We are already at a disadvantage in recruiting faculty who specialize in human embryonic stem-cell research."

9 August 2001, the date the federal restrictions came into effect. The University of Michigan uses 11 of these federally approved lines, which were grown using

mouse cells and which, they say, are less than ideal for human clinical research.

If federal law changes after the election but state law does not, Rapundalo says, stem-cell scientists in the state will be at a distinct disadvantage. Many fear that leading researchers will leave Michigan for other states that support the work, such as California and New York. Sue O'Shea, director of the Michigan Center for Human Embryonic Stem Cell Research in Ann Arbor, says she has seen many of her best students leave — especially those who want to work on embryonic rather than adult stem cells.

Morrison agrees. "We have already been at a disadvantage in recruiting faculty members who specialize in the area of human embryonic stem-cell research," he says. "If the proposal does not pass, this will not improve."

Nevertheless, unlike California's \$3-billion stem-cell agency, which was created through a ballot initiative, the Michigan initiative has no money attached to it. "So even if it does pass, it won't necessarily allow us to develop any new embryonic stem-cell lines," O'Shea says. "It will just make life easier knowing we can do it." ■

Ashley Yeager

See Editorial, page 1149.

Alzheimer's tests under fire

Genetic testing for Alzheimer's disease tells a cautionary tale about the legal, medical and ethical complications of personal genomics, as the story of a Pennsylvania company shows.

Smart Genetics, based in Philadelphia, has stopped offering its controversial 'Alzheimer's Mirror' genetic test just eight months after introducing it. The test checked for variants in a gene, called *APOE*, that bestow as much as a 15-fold increased risk of developing Alzheimer's. Soon after launching the test, though, Smart Genetics chief executive Julian Awad found himself in a controversy over whether it violated intellectual-property agreements covering *APOE* testing.

Smart Genetics' tests were performed by Athena Diagnostics, based in Worcester, Massachusetts. Athena had, in turn, licensed the patents from Duke University in Durham, North Carolina, where researcher Allen Roses discovered the *APOE* link to Alzheimer's in the early 1990s. Roses and Duke argue that Athena's licence covers *APOE* testing only in people who already have symptoms of dementia.

"The test was never intended to be used for wholesale screening of non-cognitively impaired individuals," adds Alan Herosian, director of corporate alliances for Duke University. He says he has contacted Athena many times in recent months to press this point.

Michael Henry, Athena's vice-president of business development, wouldn't comment on whether the company agreed with this interpretation of its licence. But Smart Genetics is no longer taking new orders for Alzheimer's Mirror. Its website says the test is currently unavailable because of "high demand". The company's phone lines have been disconnected, and a Philadelphia newspaper, the *Philadelphia Business Journal*, reported earlier this month that the company has closed.

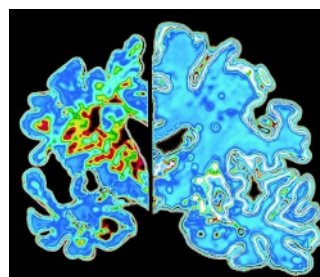
Smart Genetics co-founder Richard Watson would not comment on the newspaper article, and Awad did not respond to e-mails or phone messages. But one member of the company's scientific advisory board, Andrew Faucett of

Emory University School of Medicine in Atlanta, Georgia, notes that the firm faced another roadblock: Smart Genetics was charging hundreds of dollars for one test, whereas other gene-scanning firms offer Alzheimer's risk assessments along with other tests. "The financial model was hard to support," he says.

The tests offered by other firms bring issues of their own. For instance, Navigenics of Redwood Shores, California, provides Alzheimer's risk assessments by testing variants of a gene called *APOC1*, which sits next to *APOE* on chromosome 19. Navigenics uses *APOC1* variants to predict *APOE* status on the basis of published reports that certain variants of the two genes are often inherited together. But *APOC1* is not a perfect proxy for *APOE*

variants. Navigenics chief operating officer Sean George says the company is switching to testing *APOE* variants directly.

Roses thinks genetic testing for the risk of Alzheimer's will only get more complex. He claims to have unpublished data suggesting that variants in another gene can be used with *APOE* variants to predict, to within



Can genes reliably predict risk for an Alzheimer's brain (left)?

5–10 years, the age at which a person will develop Alzheimer's disease.

But he also hopes that ethical aspects of risk assessments will change if clinical trials identify drugs to treat the disease. Currently, knowing one's risk of developing the disease may simply cause needless worry, as there is no prevention or treatment. But Roses notes that the firm for which he worked previously, GlaxoSmithKline based near London, has reported preliminary data that its drug rosiglitazone benefits patients with Alzheimer's if they don't carry a high-risk *APOE* variant (A. D. Roses *Alzheimer's Dementia* 4, 164–166; 2008). The company is scheduled to finish clinical trials next year to test whether the findings hold up in larger numbers of patients with Alzheimer's.

If they do, Roses notes, it will mean *APOE* tests might be more useful, as they could help identify patients who will benefit most from treatment: "That's where we are approaching very, very rapidly," he says. ■ Erika Check Hayden

A. PASIEKA/SPL

No burial for 10,000-year-old bones

SAN DIEGO

In the latest twist in the tug-of-war between Native Americans and anthropologists, officials at the University of California have decided not to repatriate a pair of well-preserved skeletons that are nearly 10,000 years old.

Archaeology students unearthed the bones in 1976 near the cliff-top home of the chancellor of the University of California, San Diego (UCSD). It may be possible to extract some of the oldest human DNA in North America from the exquisitely preserved remains, say researchers. But in the past two years the bones have become a political football over US\$7-million plans to demolish and rebuild the house.

A group of 13 local bands, known as the Kumeyaay tribes, argued that the site was a sacred burial site, and that the bones found there should be repatriated to them. In March this year, UCSD dropped plans to knock down the house, opting instead for a renovation. But last week, University of California officials notified federal authorities that the bones could not be proved to be culturally affiliated with the Kumeyaay and thus would not be returned.

Steve Banegas, a tribal spokesman for the Kumeyaay, says they hadn't been notified of the decision. "They are our relatives," he says.

"We want them reburied. They should stop playing politics with the remains."

The dispute reflects the increasingly acrimonious debate over decisions involving ancient skeletons. In 2004, a federal court ruled that the roughly 9,300-year-old Kennewick Man skeleton, found in a riverbank in Washington state, should not be returned to local tribes that could not prove cultural affiliation. In other cases, usually involving younger bones, museums have returned specimens when they were shown to be culturally affiliated to local tribes.

In San Diego, tribes newly enriched by casino earnings have enlisted powerful state legislators to their cause. Facing such pressure, University of California officials are reviewing the 10-campus university's policy on how cultural affiliation is determined.

Currently, decisions about cultural affiliation are made by a panel of scientists — typically including a Native American — at each campus. Campus actions are then reviewed by a nine-person University of California panel, which includes two Native Americans, before a final decision is reached. But in September, the office of Mark Yudof, the president of the

University of California, initiated discussions about possibly eliminating the system-wide committee.

Four prominent University of California anthropologists wrote a letter to Yudof on 30 September, strenuously objecting to the proposed change. They include Phillip Walker and Michael Glassow of the University of California, Santa Barbara; Robert Bettinger of the University of California, Davis; and Philip

Wilke of the University of California, Riverside. "It is counterproductive to devolve final decision-making authority to the often inexperienced and legally ill-informed level of the local campus," says the letter.

In an interview, Bettinger said that the system-wide panel serves as a vital form of peer review. "If the analysis is not rigorous, something is missed or a campus drops the ball, the University of California system-wide panel can correct that," he says. "This has happened a bunch of times." For instance, in 2001 the system-wide committee overruled a decision by the University of California, Los Angeles, in which skeletons and funerary objects were recommended for repatriation to the Kumeyaay.

"There are many points of view on the cultural affiliation of the remains."

Gabon centre refocuses on emerging diseases

The International Medical Research Centre (CIRMF) in the former French colony of Gabon wants to have something to celebrate on its 40th birthday next year. Many hope that the centre's newly appointed director-general, Jean-Paul Gonzalez, can lead the institution — which fell into disarray in the 1990s — to a position of international excellence in global health research.

Gonzalez, from France's Research Institute for Development, will be charged with guiding the CIRMF's 157 staff and US\$5-million annual research budget. The centre, based in Franceville in southeast Gabon, boasts research infrastructure that is rare in Africa, including a biosafety level 4 lab for working with the world's most dangerous pathogens, and has long been seen as having the potential to make



Ebola virus, carried by bats, is one of the pathogens studied at the CIRMF.

a worldwide impact. It is funded largely by the French oil giant Total, which helped bring El Hadj Omar Bongo to power as Gabon's president in 1967. The centre was created jointly by the government and Total in 1979, as part of a deal

giving France access to Gabon's oilfields.

But despite its healthy finances, the centre has largely been a white elephant, some say. "The CIRMF is a very well-funded institution that has not obviously delivered

in terms of establishing a regional or international profile," adds one senior researcher in tropical disease at a leading African research centre, who asked not to be named. At the same time, he says, he understands all too well the difficulties and time involved in building up a field research centre in Africa. "The new director has plenty of scope for raising its profile and relevance," he explains.

Gonzalez took over on 1 September from Philippe Blot, who has been interim head since 2003. Blot was brought in with a remit to clean up after a turbulent period in the 1990s, when local management and expatriate researchers clashed repeatedly over administrative and other matters¹. The situation persisted until around 2003–04, and was not helped by a constant staff turnover, including

P. SALLA/SPL



California's Kumeyaay tribes are fighting to reclaim the skeletons.

Bettinger says the review ensures that science, not politics, is paramount. "The way to avoid the inevitable local politics," he says, "is to kick it to a higher level for independent review."

The turbulent history of the bones reflects this. The skeletons — of a man and a woman buried in alignment, along with the less-well-preserved remains of a third individual — were stored at the independent San Diego Museum of Man. They were also examined at the Smithsonian Institution in Washington DC. Most recently they have been at San Diego State University, where Native American scientists have been conducting morphometric analyses.

Isotopic analysis of the bones suggests that the people ate mainly seafood; anthropologists say this indicates they were seafarers, not inland dwellers like the Kumeyaay. Still, in May, the office of UCSD chancellor Marye Anne Fox wrote to Yudof's office outside the normal channels, saying that UCSD executives wanted his office to facilitate returning the skeletons to the tribes to avoid any "cultural insensitivity".

"There are many points of view on the cultural affiliation of the remains," says Art Ellis, the vice-chancellor for research at UCSD. "We wanted to make sure the office of the University of California president took into consideration all points of view when making the final decision."

Rex Dalton

three director-generals in almost as many years, and a corruption scandal.

With Blot's efforts, the worst is now over and a culture of nepotism has been cleaned up, says Patrice Debré, head of the CIRMF's scientific board. But with all the past focus on administrative matters, the centre has not had a coherent research strategy, he says. "Research fell behind; there was no real scientific leadership. That's the leadership that Gonzalez's appointment will now bring."

Gonzalez wants the centre to focus on emerging infectious diseases, which is his own research background, and to become an international centre for research into Ebola virus, arboviruses and other novel pathogens. New labs set up at the centre over the past five years have notched up some success in this area, notably the discoveries that bats

are a reservoir of both Ebola² and Marburg³ viruses. The centre has also strengthened its retrovirology research.

"We are now in an ascendant phase for research, after a long period where that wasn't the case," says Gonzalez. He hopes to reinstate parasitology and malaria research, areas that were lost by staff departures. He also wants to add a new theme: how ecosystem biodiversity affects the emergence and transmission of disease. And, to complement the centre's existing tropical-forest stations, he plans to build a field station in the savannah in southern Gabon.

The common thread running through this strategy is to refocus the centre on its key strengths. Its location in an emerging-diseases hotspot makes it ideal for studying the emergence and transmission of diseases from primates to humans.

Furthermore, the CIRMF's primate centre is among the

largest in Africa. It contains around 450 primates, including gorillas, mandrills and macaques, half of them housed in a forest enclosure.

"The natural hosts for SIV [simian immunodeficiency virus] and STLV [simian T-lymphotropic virus] and many other viruses are African primates," says Bettina Salle, the veterinary surgeon who runs the primate centre.

"We have the ideal animal colonies and geographical location for research into the viruses in their natural hosts."

Debré notes that the centre can't begin to compete with international clinical research in retroviruses such as HIV, in part because it lies too far from major hospitals. So instead it hopes to capitalize on the current back-to-basics shift in AIDS research, to position itself as a place to study wild retroviruses in primates and carry out HIV research and vaccine trials in primate models.

The CIRMF should take advantage of the infrastructure provided by the government and by Total, says Debré, but balance that by pursuing outside, independent research funding. Gonzalez says he hopes to have obtained a 50% increase in the centre's research budget and staff by 2010. "A test of Gonzalez will be how successful he is here," says Debré.

One former CIRMF researcher praises Gonzalez's track record, but reckons the director-general faces an uphill battle. Salle is more optimistic: logistics and cultural challenges mean that working at the CIRMF is "not always easy", she says, "but we can do great things here. Gabon has a great tool for science."

Declan Butler

1. Butler, D. *Nature* **394**, 113 (1998).
2. Leroy, E. M. et al. *Nature* **438**, 575–576 (2005).
3. Towner, J. S. et al. *PLoS ONE* **2**, e764 (2007).



HAVE YOUR SAY
Comment on any of our
news stories, online.
www.nature.com/news

German authority halts primate work

Germany's constitution guarantees its citizens the freedom to conduct research — but local authorities in the northern city of Bremen are forcing a leading neuroscientist to halt his primate experiments. A court will probably now have to decide whether the controversial ruling violates federal law.

Andreas Kreiter at the University of Bremen uses 24 macaques to study cognitive processes in the mammalian brain. Germany's largest animal-protection group, the Animal Welfare Association, has for years campaigned against the experiments, claiming that they are intolerably painful and have no short-term therapeutic use.

Local politicians have become increasingly sympathetic to that view. Last year, in a move criticized by scientists as a grab for votes, Bremen's parliament called on the state government to ban Kreiter's primate research (see *Nature* 446, 955; 2007). After regional elections in May 2007, the newly formed Social Democrat–Green coalition government agreed not to reapprove his experiments when his current licence expires later this year.

On 15 October, Kreiter was officially informed by the senate of health — the local authority in charge of approving animal experiments — that his licence will not be renewed. Referring to “changed societal values”, the authority argued that the experiments were “ethically unjustified” because they address long-term scientific questions rather than help develop specific medical therapies.

Kreiter counters that the authority's reasoning is subjective and will not hold legally. “It is a purely arbitrary ethical assessment that doesn't go into the nature and purpose of my work at all,” he says.

The authority's logic is “pretty strange”, agrees Reinhard Fischer, who oversees the primate experiments on behalf of Wilfried Müller, the university rector. Fischer notes that the ruling ignores a positive judgement rendered last year by an expert commission comprising scientists and representatives of animal-welfare organizations. Bremen's government had asked the commission to evaluate Kreiter's work, and it voted in favour of continuing the experiments — deeming them scientifically important and in full accordance with animal-protection laws.

Matthias Kleiner, the president of Germany's main research-funding agency, the DFG, and many neuroscientists have criticized the



Andreas Kreiter plans to appeal the controversial decision to stop his studies of cognition in macaques.

decision. “We're all very concerned about what's going on in Bremen,” says Stefan Treue, director of the German Primate Center in Göttingen and coordinator of EUPRIM-Net, a European Union-funded network of eight large primate centres. “I just can't see why what's perfectly

fine in one place should be unethical in another. To me, this is a flagrant case of political interference with approval procedures, and with the general freedom to do science.”

Nine laboratories across Germany conduct primate experiments; in recent years

the number of animals involved has ranged between 1,300 and 1,800. About 150 of those are for basic research; the rest are used for other experiments such as drug toxicity tests.

Kreiter's experiments record the activity of single neurons in the brains of macaques while they complete particular behavioural tasks. This cannot be usefully done on animals with less-developed brains, such as mice, or with available non-invasive brain-imaging

methods, says Kevan Martin, a neuroscientist at the Swiss Federal Institute of Technology, Zurich, whose own research on macaques is also threatened by a ban (see *Nature* 453, 833; 2008). “By taking away monkey experiments you certainly greatly diminish the potential benefits of fundamental research like Kreiter's for future clinical applications,” says Martin.

The ruling jeopardizes the existence of Kreiter's centre, including the work of a dozen or so postdoctoral researchers and PhD students. Also at risk is his group's participation in a €2.3-million (US\$2.9-million) project, funded by the German science ministry and aimed at developing key technologies for neuroprostheses.

Kreiter's current licence to conduct experiments runs out on 30 November. To be able to continue his research beyond that date he will appeal the decision and seek an interim injunction at court.

University officials say they will support him and, if necessary, take the case to Germany's Federal Constitutional Court.

Quirin Schiermeier

J. SARBACH/AP

Salmon study sparks row over dams

Research that argues dams have no direct effect on the migration of juvenile salmon is roiling waters in the US Pacific Northwest.

The study, published this week in *PLoS Biology*, uses a new way to tag fish to compare the heavily dammed Columbia River system with the free-flowing Fraser River to the north in British Columbia, Canada. After hatching upriver, salmon smolts fared equally well on their journeys down the two rivers to the ocean. But that's where the agreement ends.

Salmon migration remains at the heart of an ongoing court battle, which pits environmentalists who want to see some of the dams removed against federal agencies charged with maintaining the dams, providing power and protecting salmon. Thirty-one federal dams on the Columbia River and its tributaries collectively provide some 60% of the region's electricity.

The Bonneville Power Administration, based in Portland, Oregon, has spent billions of dollars making the dams more friendly for fish, and the latest plan, issued on 5 May, proposes further modifications to dam operations. Taken at face value, the new research might suggest that the investment is paying off and that other factors are suppressing the 13 salmon populations protected under the Endangered Species Act.

But environmentalists and several salmon biologists pounced on the study, suggesting that industry funding might have biased the results. These critics question the value of comparing the two rivers and say that the study doesn't even address what many think is the dams' biggest effect: stressed smolts dying after they reach the ocean.

Still, the study explicitly acknowledges these limitations. Lead author David Welch, founder of Kintama Research in Nanaimo, British Columbia, and an adjunct professor at the University of Victoria, has received nearly US\$4.2 million from Bonneville Power for his work, although his largest funder is the Gordon and Betty Moore Foundation. However, he says that his funders had no role in vetting the study. "Everybody assumed that we would simply see much higher mortality in the dammed system — we didn't find that," he says. "We don't want to oversell it and say it definitively answers these questions, but it's pretty damn surprising."

But two of Welch's co-authors, Shaun Clements and Carl Schreck, who did their work at Oregon State University in Corvallis, are concerned that the study is being oversold.



M. CONLIN/V&W/IMAGEQUESTMARINE.COM

Free-flowing water might not aid chinook salmon as much as some believe.

Clements says that a press release issued by *PLoS Biology* with the headline "Dams make no damn difference to salmon survival" goes too far. "This study doesn't take dams off the hook," he says. It may point to problems in the Fraser rather than clearing dams on the Columbia, he adds.

Sound trackers

Tagging studies typically rely on passive sensors implanted in the fish, but the sensors have a short range and can be counted only as the salmon are funnelled past receivers at the dams. In contrast, Welch's company uses powered transmitters that send acoustic signals to receivers, allowing the researchers to track the salmon in free-flowing rivers and along the coast.

The team tagged hatchery spring chinook and wild steelhead on the Fraser River from 2004–06 and then compared

those results to other tagging work done in the Columbia. In 2006, the team also tagged hatchery spring chinook fish on the Columbia's Snake River tributary and tracked fish as they navigated eight major dams to the ocean. The numbers of surviving fish varied widely, Welch says, but were statistically indistinguishable between the river systems (D. W. Welch *et al.* *PLoS Biol.* 6, e265; 2008).

Michele DeHart, manager of the Fish Passage

Center in Portland, which tracks salmon statistics, says the comparison is "seriously misleading." She points out that the Fraser River has its own problems, including an infestation of bark beetles, which could affect water quality, and that federal dam operators have been under a court order since 2005 to help juvenile salmon by spilling water over the dams during spring and summer migrations.

Ed Bowles, the top fish biologist for the Oregon Department of Fish and Wildlife, says the results merely underscore the notion that delayed mortality is the real killer. He points to work showing that the number of salmon returning upriver to their original spawning grounds — years after their initial migration to the ocean — is several times higher for fish that navigate fewer dams in the lower reaches than for populations that must traverse eight dams to reach the sea (H. A. Schaller and C. E. Petrosky *North Am. J. Fish. Mgmt* 27, 810–824; 2007).

The additional energy and stress required to move through the dammed system seems to inhibit survival in the ocean, he says. "This paper doesn't really do anything to dispel that working hypothesis."

Welch says he is now looking into the issue of delayed mortality by tracking the same salmon populations after they enter the ocean and head up the coast to Alaska.

Jeff Tollefson

"Everybody assumed that we would simply see much higher mortality in the dammed system."



MEDICAL ISOTOPES
European shortage will
continue into 2009.
www.nature.com/news

Older scientists publish more papers

If the mind is the first to go, it doesn't seem to show up in the bibliography. A new study suggests that older scientists publish more than their younger counterparts.

A research team from the University of Quebec in Montreal, Canada, looked at the publication record of nearly 14,000 professors and found that those in their 50s and 60s published almost twice as many papers each year as those in their early 30s. What's more, the papers are just as highly cited, according to the work, which appears on the arXiv preprint server (Y. Gingras *et al.* <http://arXiv.org/abs/0810.4292>; 2008).

The survey, say its authors, is the strongest evidence yet that older scientists are at least as productive as their juniors, if not more productive. "We should not underestimate the older researcher," says team leader Yves Gingras, a historian of science.

Gingras became interested in the topic after

reading that the average age for first grants from the US National Institutes of Health in Bethesda, Maryland, rose from 34 in 1970 to nearly 42 in 2004. His team looked at publication data for 13,680 scientists and scholars in Quebec between 2000 and 2007, noted each researcher's age, the number of papers published and the citation record according to the Thomson Reuters citation index.

The findings held across both the sciences and the humanities. Unsurprisingly, the team found that productivity increased sharply throughout the late 20s and early 30s. But it continued to rise at a

slower rate for researchers in their 40s and 50s. Active professors in their 60s continued to produce about three papers per year. The average impact factor of each paper fell during the rise in productivity, but turned upward again for papers published by older authors. Citations, meanwhile, remained stable, suggesting that

older researchers were both more prolific and more highly cited than their juniors.

The survey is one of the largest of its kind, says Wolfgang Glänzel, a director at the Policy Research Centre for R&D Indicators at the Dutch-speaking Catholic University in Leuven, Belgium. He would like to know how many of the papers published by older authors are review articles, and what effect students had on their older co-authors' output. Both could be important factors in the apparent high productivity of senior scientists, he says.

"High impact is not the same as originality," adds Anthony van Raan, who studies science and technology policy at Leiden University in the Netherlands. Still, he says, the research makes a compelling case against mandatory retirement, which is commonplace throughout Europe. "It's better to keep [researchers] and let them work," he says. "They can be very productive." ■

Geoff Brumfiel

"It's better to keep researchers and let them work."

Rare greenhouse gas adds to climate concerns

Scientists are recommending that a rare but potent greenhouse gas should be included in future climate agreements after confirming that the gas is about four times more abundant than previously believed.

Nitrogen trifluoride (NF₃) is commonly used to etch microcircuits in plasma-screen televisions and other flat-panel displays. As a greenhouse gas, NF₃ is about 17,000 times more efficient at trapping heat than carbon dioxide, but very little was thought to escape into the atmosphere.

A team led by Ray Weiss of the Scripps Institution of Oceanography in La Jolla, California, confirmed that atmospheric concentrations have risen more than 20-fold during the past 30 years after analysing air samples from coastal stations in California and in Tasmania, Australia.

The scientists reported that in 2008 some 5,400 tonnes of NF₃ were present in the atmosphere (R. F. Weiss *et al. Geophys. Res. Lett.* doi:10.1029/2008GL035913; 2008).

For a longer version of this story, see <http://tinyurl.com/6a6gta>

Rice pathogen is added to list of bioterror agents

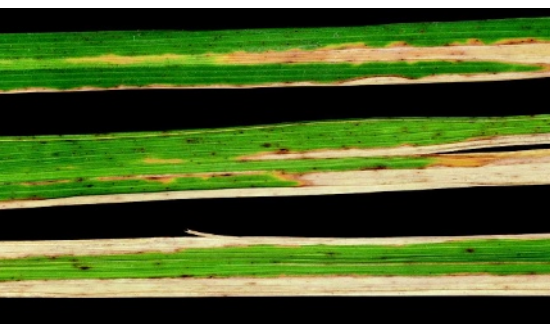
The US Department of Agriculture (USDA) this month added a rice pathogen to its national security watch list, despite objections from several prominent scientists.

USDA's Animal and Plant Health Inspection Service has listed the bacterium *Xanthomonas oryzae* pv. *oryzae*, which causes leaf blight, as a 'select agent' that could be used for bioterrorism.

Several researchers who work with the microbe argued against the listing, saying that it would have difficulty establishing itself in the United States, and that those labs working with it have perfect safety records dating back more than 20 years.

Researchers have until 17 November to notify the government if they possess the pathogen, and until 14 April 2009 to come into full compliance with the regulations.

"That's not a lot of notice," says Pamela



Blight in rice caused by *Xanthomonas oryzae*.

India makes history with launch of Moon mission

India's first lunar spacecraft is set to swing into orbit round the Moon on 8 November.

The Indian Space Research Organisation (ISRO) successfully launched Chandrayaan-1 (pictured) on 22 October from the Satish Dhawan Space Centre in Sriharikota, southern India.

The craft is carrying six foreign and five Indian scientific instruments designed to create a three-dimensional map of the Moon and study the lunar soil, among other things. The spacecraft will also search for frozen water and measure the abundance of helium-3.

The ISRO is now planning Chandrayaan-2, which would take a Russian-built rover to the Moon's surface by 2012. The agency also aims to send two Indians into space by 2015, in a mission that will cost roughly 120 billion rupees (US\$2.4 billion).



ISRO/AP

Ronald, a rice researcher at University of California, Davis, who opposed the listing. "All the labs are scrambling to find out what it means to their research programmes."

Fall in profits prompts major job cuts at Merck

Pharmaceutical giant Merck is to cut 7,200 jobs, or roughly 12% of its workforce. Around 40% of the cuts will be in the United States, many of them from management as Merck trims the number of senior and mid-level executives by 25%. The company, based in New Jersey, also plans to close basic-research sites in Tsukuba (Japan), Pomezia (Italy) and Seattle (Washington).

These cuts come on top of the axing of 10,400 positions announced in 2005, which Merck says was "substantially complete" as of last month.

Merck's third-quarter profits were down 28%, and spending on research and development (R&D) dropped by 19% compared with 2007, to \$1.2 billion. But the company says that R&D spending actually rose by 2% if expenses from restructuring efforts and the 2007 acquisition of NovaCardia are taken into account.

Swedish basic research receives funding boost

The Swedish government has announced a record investment in basic research, much of which will be directed to fields such as stem cells, climate modelling and nanotechnology.

Introduced on 23 October, the bill provides almost 15 billion Swedish krona

(US\$1.9 billion) for research and innovation between 2009 and 2012. Sweden currently spends 25 billion krona a year on research; that figure is expected to rise by 20% by 2012.

Most of the new funds will be distributed to universities through a merit-based system. Gunnar Öquist, head of the Royal Swedish Academy of Sciences in Stockholm, says that although the initiative would reverse years of declining research budgets, he worries that its narrow focus could stifle creativity.

The bill will now be taken up by the parliament, and is expected to pass early in 2009.

NASA clamps down on conference attendance

NASA has put a moratorium on participation in conferences in an effort to comply with an authorization bill recently enacted by Congress.

Signed into law by President Bush on 15 October, the act limits conference spending to US\$5 million in the 2009 fiscal year. That figure represents more than a two-thirds reduction from conference spending in 2008.

The moratorium, announced earlier this month, covers sponsorship of conferences, as well as travel and admission fees. It applies to NASA staff and some contractors, such as employees at the Jet Propulsion Laboratory in Pasadena, California. Scientists receiving NASA grants or sponsorship have not been affected so far.

Correction

The News Feature 'What lies beneath' (*Nature* **455**, 724-725; 2008) incorrectly stated that nitrogen-fixing bacteria are 1-2 millimetres in length. They are, in fact, 1-2 micrometres long.



Where should the drug industry go to find new ideas? In the first of two features, **Alison Abbott** asks if the future lies in systems biology — a field that attempts to piece together ‘everything’. In the second, **David Cyranoski** looks at drug companies’ attraction to China.

A FIENDISH PUZZLE

If it is hard to make something from nothing, then it can be just as hard to make something from everything. But that, in essence, is what many pharmaceutical companies are trying to do as they seek new drug targets by integrating the massive sprawl of biological information now available.

The puzzle is like a jigsaw with an unknown number of pieces and, as yet, no edges. So when Cristiano Migliorini of the drug giant Roche saw a promising drug target for type 2 diabetes emerge from hundreds of millions of data points, he felt a celebration was in order. He picked up the phone and invited his 20 or so collaborators from research institutions all over Switzerland to a party.

The celebration — a good dinner, drinks and some lively pub games — might have been considered premature. It is a long and tortuous route from an interesting protein to a new and effective drug. But since the party, this protein has been put through its paces at Roche’s drug-development labs in Basel, and candidate drugs that bind to it are already lined up for testing. The target could still fail, but in Migliorini’s mind, the results have already shown that the mega-data-crunching approach of ‘systems biology’ — new, modish and increasingly adopted by pharmaceutical firms — can piece together a meaningful picture from this colossal biological puzzle.

“And that,” says Migliorini, “is worth at least a good meal and a game of pool”.

The crisis-ridden pharmaceutical industry desperately wants something to celebrate. The glory days of the blockbuster drug seem to be over. When genomics matured at the turn of the century, much of the industry was convinced that individual genes would emerge as the new drug targets. But that reductionist bubble soon burst: the more that geneticists and molecular biologists have discovered, the more complicated most diseases have become. As individual genes have fallen out of favour, ‘systems’ — multitudes of genes, proteins and other molecules interacting in an almost infinite number of ways — have come into vogue. Systems biology is an attempt to make sense of all these data.

Some researchers and analysts were cynical when systems biology was hyped as the saviour of the failing research and development pipelines. “Rightly so, perhaps,” says Giulio Superti-Furga, who studies complex protein interactions and is head of the Research Center for Molecular Medicine in Vienna. “A few years ago, systems-biology proponents were writing reviews that triggered unreasonable expectations for advances in medicine.” Many companies have therefore been cautious about making a big investment. Like Roche, most are testing the water with defined projects, usually

in collaboration with top academics or small computational companies.

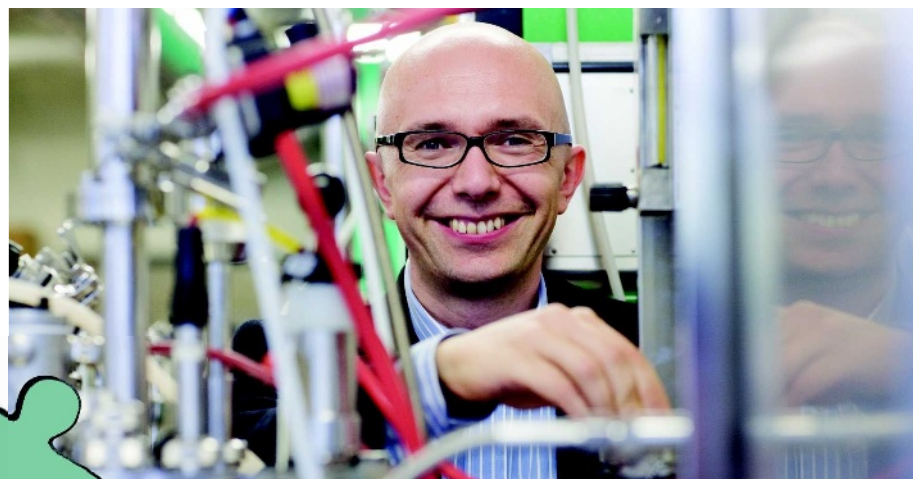
Although systems biology is far from proven as a drug-discovery tool, it has many advocates and its wider adoption seems to be inevitable. If leading biologists are concluding that disease can only be understood by tackling the entire, messed-up system, pharmaceutical companies arguably have little choice but to keep up. “Industry really needs to dive in,” Superti-Furga says.

Never-ending networks

At its heart, systems biology is about gathering unprecedented amounts of data from cells, then making sense of it through mathematical models. At its most sophisticated, it might involve the high-throughput collection of molecular data, such as DNA sequences, RNA molecules, proteins and metabolites as well as more descriptive data such as clinical diagnoses and drug responses. These data are then assimilated into computational models of cellular processes which, as molecules change location and function every microsecond, must also accommodate the dimensions of time and space.

It is not enough to mimic what’s known about the cell — such models must also predict what is unknown so that scientists can test their hypotheses. A model of a liver cell, for example, might reveal which of 100 interesting drug candidates will prove beneficial and which will be toxic. Knowing when to pull a doomed drug from the pipeline could spare a company a lot of expensive work. “The biggest value of systems biology immediately is as an aid to decision-making,” says David de Graaf, who has worked in systems biology programmes at Pfizer, AstraZeneca and now at Boehringer Ingelheim. “And a drug company is a decision-making machine.”

Although systems biologists want it all, for now they cannot get it. “We’ll not have everything a hundred years from now,” says Colin Hill, chief executive of the biotech company Gene Network Sciences in Cambridge, Massachusetts, which uses supercomputers to model complex biology for drug companies. “Even after two decades of gathering and



Cristiano Migliorini has seen systems biology deliver promising drug targets.

ILLUSTRATION: J. HEIN VAN DIEN DONCK
C. ROTH

J. RADIN

archiving molecular data, barely 5% of the total circuitry inside the cell is known." So systems biologists must choose between two approaches when developing their models. In one, they scale back their ambition, paring down the number of biochemical pathways being modelled to just a few well-known ones. In the second, they bluff the bits they do not know, trying to model 'everything', and using sophisticated computational tricks to compensate for absent data.

At the end of the day, the choice of approach depends on what question researchers want to ask of their models. It's a bit like the conundrum in Douglas Adams's *Hitchhiker's Guide to the Galaxy*, where the computer Deep Thought spends 7.5 million years computing an answer to the most horribly imprecise "ultimate question of life, the Universe and everything" — only to come up with the answer "42". If the question is narrow — 'which surface proteins are affected when insulin activates receptors on the pancreas cell?' — then it may be sufficient to model only those networks known to be involved in insulin signalling. If the question is broad — 'what would happen to an animal fed drug X?' — then the 'everything' approach is more appropriate to see what falls out, be it stronger muscles or liver failure.

The meaning of life

Drug companies are divided on which of the two approaches is better, and Migliorini is clear about which he prefers. "We take modelling in small doses — a defined number of well-understood pathways," he says. "We have a pragmatic approach — we can't say '42' to management." Cell biologist Willy Krek, an academic partner of Migliorini's at one of Switzerland's foremost universities, the ETH in Zurich, adds: "It is strategically better to get a rich data set on a limited network than to take measurements for the rest of your life, always saying they are still not enough."

Even the scaled-back approach to systems biology takes massive effort. Migliorini's collaboration began in 2006 as part of Switzerland's big systems-biology research programme now known as SystemsX.ch. Migliorini calls it "an industry-academic collaboration beyond cliché", referring to what is in fact the largest such partnership between pharmaceutical companies and academia in biology that Switzerland has ever seen. It involves several top Swiss research groups, plus academic and industrial partners elsewhere, at a cost of 9 million Swiss



Colin Hill uses his background in theoretical physics to model complex biological systems.

francs (US\$8 million) over three years as well as undisclosed internal spending.

In searching for its diabetes drug, the consortium has built a model focused on the 500 or so proteins — called the 'surface proteome' — on the coat of pancreatic β cells, the cells that secrete insulin when blood glucose concentrations rise. Migliorini and his co-workers are working on the concept that some people who overeat develop type 2 diabetes because their β cells cannot keep up with the body's demands. The team is seeking to identify surface proteins that stop responding to environmental signals such as glucose metabolites, as these may be responsible for transforming the overworked cells into diseased cells.

The scientists use almost every high-tech method imaginable to gather data on proteomes and the cellular networks they link into. Consortium member Rudi Aebersold from the ETH developed one such method using mass spectrometry that separates out glycoproteins — the proteins involved in key processes such as protein folding or cell-cell signalling — from the background of other proteins¹. Importantly, the technique quantifies as well as identifies proteins, providing a molecular fingerprint for a cell in a particular stage of health or disease. Quantification is essential to systems biology, whereas in the past, molecular biologists were happy just to see whether or not a protein was present.

To flesh out the proteomic information, the researchers also generate reams of other data, either experimentally or by pulling them out from public and commercial databases, to feed into their computational model. They then work out which surface proteins converge on the same intracellular pathways. At the end, they have a massive, moving, knitting-ball of a network in which every protein is linked to every other protein it interacts with. The hope is that this

can point to the 'nodes' — the key points in a network that have disproportionate influence.

Then researchers can ask questions of the model. For example, if the team suspects that the disease is being caused by a particular aberrant glycoprotein, they strip that protein from the model. If correct, the model should then mimic the knock-on effects in the circuitry that are known to occur. That glycoprotein then becomes a potential drug target or, perhaps, a biological marker of disease progression.

Quick results

At the start, not everyone at Roche expected the systems approach to succeed within the three years the collaboration was initially given. In fact, Migliorini's first hit, last autumn, came within a year of his project starting. "It was a great feeling to see our engineering-style approach — multiple inputs, each comprising amounts of data that make you sweat — spit out a clear single target," he says. Here, luck intervened: the protein in question (its identity is still guarded) had been investigated in relation to an entirely unrelated disease, so the tools and compounds for studying it were already available. "The whole process went lightning-fast," says René Imhof, director of research for Roche in Basel. Roche's academic partners are also happy, because the collaboration was formed at exactly the time they were starting to feel the limitations of the traditional genes-will-explain-disease reductionism.

Across the Atlantic, in the Boston area of Massachusetts, Peter Sorger and Douglas Lauffenburger have been feeling similar constraints. Sorger says that he used to regularly raise hackles with his outspoken criticism of traditional genomics. As the region is home to the Whitehead Institute for Biomedical Research, one of the most important sequencing hubs of the Human Genome Project, those hackles were being raised on some particularly high-level necks.

Sorger and Lauffenburger now operate complementary systems-biology labs — Sorger at Harvard Medical School in Boston and Lauffenburger at the Massachusetts Institute of Technology in Cambridge — and have been asked to collaborate with numerous pharmaceutical companies, usually in the area of cancer. Both have worked with AstraZeneca, for instance, to understand the action of the company's anticancer drug gefitinib (Iressa). Gefitinib was a much-heralded drug designed to hit the ERBB1 receptor when it is mutated, as it frequently is in various cancers. But few patients responded to the drug, and genetic sequencing showed that mutations in the

"We'll not have everything a hundred years from now."
— Colin Hill

receptor could not predict who these patients would be. "By focusing down on just the gene, the genomic approach had taken the target out of its physiological context," says Sorger, "and the physiology needed to be put back with systems biology."

Lauffenburger attempted to do this by looking at the whole life cycle of the errant receptor and how it connected with biochemical pathways inside the cell. Like many membrane receptors, ERBB1 switches off its signalling capacity by removing itself from the surface of the cell, and Lauffenburger theorized that sensitivity to gefitinib could depend on how well this process works. His team collected data on genes, transcription, protein phosphorylation and the internalization of the receptor from cells that were either sensitive or insensitive to gefitinib and built a model to mimic the circuitry associated with the receptor.

The model suggested that mutations in ERBB1 were almost irrelevant: gefitinib would work only in cells that had inefficient internalization of ERBB1 (ref. 2). So far this has not generated new biomarkers for the company's clinical trials, but it has shown that complicated systems biology may be needed to select patients for drug treatments.

Like Migliorini and many others in systems biology, Lauffenburger has an engineering background. Others, such as Hill, have swept in from theoretical physics. Hill turned to biology after a stint at the interdisciplinary Santa Fe Institute in New Mexico, where he realized that physicists like himself had become comfortable contemplating the vastness of the Universe or the minuscule nature of the subatomic world, but were distinctly uncomfortable with a mid-sized affair such as life. "Life was always too complicated a subject for us," he recalls.

Hill's type have brought with them the sort of Bayesian, or probabilistic, analyses that have been applied in other areas that have huge and incomplete data sets, such as astronomy. They use these techniques to tackle the second approach to systems biology, relying on computer models to fill in the gaps between relatively sparse data points. Researchers then put real data into the model and modify it to fit the data better, repeating these adjustments until all possible permutations have been explored.

This type of modelling requires a scale of computing effort analogous to that required to predict weather and understand global warm-

ing. The number of ways in which one gene or protein can link to another in the tangled web of intracellular pathways and networks is astronomical. "If you take 22 variables, such as gene, transcript or clinical measurements, they may each affect an average of maybe 250 points in a network — that means there could be 250²² possible combinations," says Hill. "We can only do this with supercomputing." Hill's company is now cranking out potential drug targets, which it patents and then peddles to pharmaceutical companies.

Industry giant Merck is known for its bold move to incorporate the 'everything' approach into its drug-development programmes. In 2001, it acquired the Seattle-based genomics company Rosetta Inpharmatics to develop systems genomics in house. "The scale of the biological information we have today is unprecedented," says Eric Schadt, Rosetta's science director and a former physicist. "Fifty thousand or so transcripts, millions of mutations, hundreds of clinical endpoints — a high-performing computing environment is essential and we have one of the world's biggest."

Growth factors

Earlier this year, the group used the model to show that its systems approach had identified key disease-causing gene networks in metabolic diseases³. The paper was published back-to-back with another that had used a systems approach to identify three obesity-related genes that had not sprung out of genomics data alone⁴. "The surprise here was to find that not just a handful of genes were causally implicated in the disease, but hundreds — whole pathways were involved," says Schadt. "So the challenge is to hit the nodes in networks where implicated pathways overlap." The company says that up to 30% of targets for

various diseases in its early pipeline have been identified through Schadt's techniques, and that a few are in early-phase clinical trials.

But some people in industrial and academic circles remain hesitant about systems biology. No one can be sure that it will really increase the number of targets or biomarkers that make it through clinical testing. Still, it is a gamble that almost all companies seem willing to make, even if investment levels are sometimes small. In a report published earlier this year, industry analysts at PriceWaterhouseCoopers argue that the pharmaceutical industry needs to rely much more on systems biology if it is to survive the failing-pipeline crisis and predicts that the approach will have become more prevalent by 2020.

Stephen Friend, Merck's vice-president for oncology, thinks that any hesitancy will be overcome when the modelling becomes so predictive that the toxicity and efficacy of a potential drug can be forecast very accurately even before an experimental animal is brought out if its cage. "The next three to five years will provide a couple such landmark predictions and wake everyone up," he says.

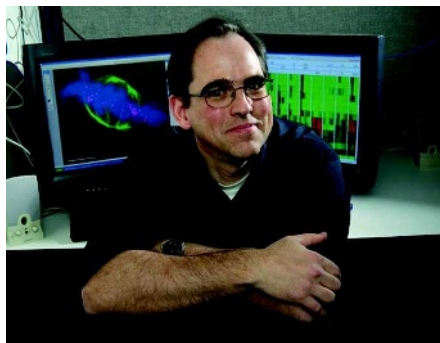
The challenge for systems biologists is that the system is being stretched at all ends of the imagination. There will inevitably be new categories of biological information to collect and feed into models. Measurements will be made on finer and finer scales so that individual molecules can be identified and counted in single cells with greater speed and efficiency. This information will be integrated into grander models, although hopefully not ones that take 7.5 million years to compute answers. At a meeting in Japan in February, scientists signed up to a declaration calling for a grand challenge to create a virtual representation of the physiology of the entire human within three decades. To get that far that fast will require formidable advances in technology development, and even greater computational power and confidence.

But for the time being, most systems biologists are happy to have their questions answered, or at least considered, by drawing boundaries around 'everything' such that it is contained within the walls of a virtual cell. "In fact, I don't think we have a choice," Migliorini says. Life, the Universe and almost everything is about the most that biologists can handle for now.

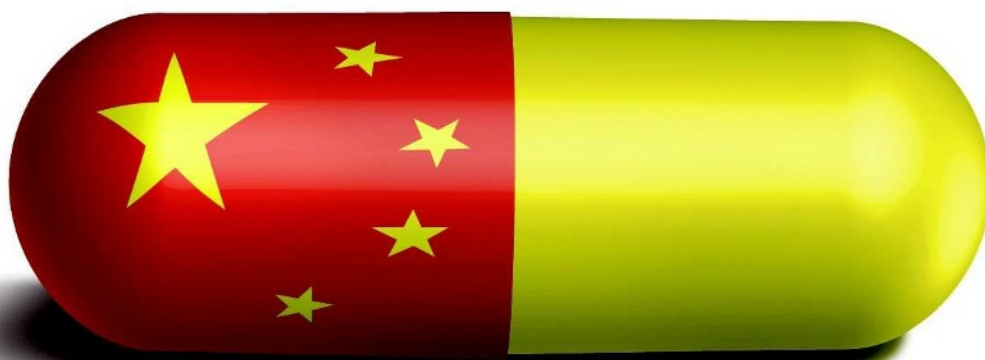
Alison Abbott is Nature's senior European correspondent.

1. Zhang, H., Li, X. J., Martin, D. B. & Aebersold, R. *Nature Biotechnol.* **21**, 660–666 (2003).
2. Hendriks, B. S. et al. *IEE Proc. Syst. Biol.* **153**, 22–33 (2006).
3. Chen, Y. et al. *Nature* **452**, 429–435 (2008).
4. Emilsson, V. et al. *Nature* **452**, 423–428 (2008).

"Physiology needed to be put back with systems biology."
— Peter Sorger



Systems biologist Peter Sorger has criticized genomics techniques.



MADE IN CHINA?

From the outside, China may not seem an obvious place to look for a pharmaceutical breakthrough. Last year, the head of the State Food and Drug Agency was executed for taking bribes. This year, more than 80 people died in the United States alone after taking a contaminated version of the blood-thinner heparin, the main ingredient for which had been imported from China. The number of domestically discovered drugs that the Chinese pharmaceutical sector has turned into internationally successful products? Zero.

But these are desperate times. Pharmaceutical companies around the world face shrinking incomes as their leading drugs go off-patent and generics come online. The cost and complexity of new drug discovery and clinical testing is soaring and drug firms are taking a machete to their research and development (R&D) programmes.

Looked at in this light, certain features of China begin to look attractive. A spate of returnees, often with experience from Western pharmaceutical companies, has put talent on the ground. A vibrant contract research organization (CRO) industry — with more than 400 companies in Shanghai alone — allows for speedy changes of direction with minimal investment. Clinical trials can cost half or even one-fifth of what they would in the United States. The government is showing that it is serious about meeting international standards for drug development with a 6-billion renminbi (US\$900 million) investment aimed at good practices in laboratory and clinical practice (see *Nature* 455, 142; 2008). And, perhaps most importantly, multinational

companies recognize that a research branch in China helps build connections with regulatory agencies and gain access to a market that some analysts predict will be the world's second largest after the United States by 2020.

The combination of desperation outside China and promise within has convinced almost every big pharmaceutical player, including Roche, Novartis, Glaxo-SmithKline, Eli Lilly and Pfizer, to collectively invest hundreds of millions of US dollars into research operations there over the past two to three years. The

companies are somewhat cagey about how much they are investing at a time when they are laying off employees elsewhere, and when there is no guarantee of a return. But Kenneth Chien,

"You can consider what we are doing in China an experiment. We don't have to own everything."

— Tony Zhang



Jingwu Zang says his firm was attracted to China by the potential workforce and drug market.

an expert in cardiovascular medicine and an adviser to several large drug and biotechnology firms working in China, calls it "Basel on steroids", referring to the throng of pharmaceutical companies in Switzerland.

Expansion into China is an experiment, allowing companies to try out new models for seeking drugs, often relying heavily on outsourcing to CROs. "Every one is doing different things. Eventually we'll figure it out," says Li Chen, head of the Roche R&D centre in Shanghai. But there is no proven leadership, points out Mingwei Wang, director of China's National Center for Drug Screening and a pharmacologist at the Shanghai Institute of Materia Medica, the drug-research institute of the Shanghai Institutes for Biological Sciences. "Pharmaceutical companies in China have all the problems of the United States plus the lack of experience. There are no success stories, nobody has built an Amgen or Genentech in China."

Hi-tech high-rises

Most of Shanghai's pharmaceutical boom is happening in the Zhangjiang Hi-Tech Park in Pudong, east of the Huangpu river. It is in the 'new' part of Shanghai, symbolized by the Oriental Pearl Tower, with wide streets, spacious new residential complexes and golf courses. Zhangjiang boasts companies selling almost every service a pharmaceutical or biotech company could want, including the production of active ingredients, genomics, analytical and combinatorial chemistry, and preclinical toxicology testing. It is a natural place for a drug company to want to set up in.

But not GlaxoSmithKline (GSK). GSK plans to go to Puxi, the area west of the Huangpu, to build a massive permanent home for its research branch, launched in May 2007. Puxi

ILLUSTRATIONS: W. FERNANDES

GSK

GSK

has *shikumen* stone-gate dwellings and fashionable restaurants, bars, and cafes tucked into alleyways. Puxi is also home to Shanghai's, and probably China's, most prized biological sciences research institute, the Shanghai Institutes for Biological Sciences, in addition to strong research hospitals, such as the Ruijin and Zhongshan hospitals, and the medical schools of Fudan University and Jiao Tong University. "We were looking for a scientific environment," says Jingwu Zang, who heads GSK's new outpost, called R&D China. "Zhangjiang is for services. We want hard-core science."

R&D China's motto is 'Discovered in China', and it is emblazoned over an image of the Oriental Pearl Tower on the wall of the company's temporary office in Zhangjiang. Zang hopes that the phrase will replace the ubiquitous 'made in China' that brands the mostly low-margin goods made there. The unit will be the global centre for GSK's new neurodegenerative and neuroinflammatory drug-development programmes, focusing on multiple sclerosis, Parkinson's disease and Alzheimer's disease. The plan is to employ 1,000 people over the next 10 years. The company declined to give the full cost of the enterprise, but such a unit would be three to four times larger than those of other drug companies investing in the country.

GSK's decision to build such a large and permanent facility, rather than rent space as most companies do, suggests great confidence in what can be accomplished in China. It is particularly striking considering that GSK announced plans last autumn to lay off 5,000 people in other branches. Zang says that the company was drawn to China by several factors, and that cost-saving is not one of them. One attraction is the growing talent pool of scientists who trained overseas and have already returned or are willing to return. Almost half of R&D China's scientists are people who have returned from working in the United States or Europe, says Zang. Researchers are being lured back by government-supported research institutes, university positions and the opportunity to start their own government-supported businesses (see 'Started in China'). He says the company also has in mind the benefits of being able to transition drugs to the growing Chinese market more easily.

Zang rates the "overall interactive scientific environment" in Puxi as comparable to that in the Boston area of Massachusetts and hopes



Most drug firms expanding to China rent offices in the Pudong district of Shanghai.

N. ELIAS/REUTERS

to thrive in it. Over the past decade, GSK has collaborated with the Shanghai Institute of Materia Medica and it plans to do more with the Shanghai Institutes for Biological Sciences and the Shanghai Institute of Organic Chemistry.

China also offers opportunities to try an Asian angle on some intractable research problems. For example, Zang plans to investigate how multiple sclerosis differs between Asian and white populations, drawing on patients from Ruijin, Zhongshan and other hospitals. Until now, most researchers studying multiple sclerosis, such as those in the Consortium of Multiple Sclerosis Centers, of which GSK is a member, have mainly studied white people. "In the United States we hit a wall," says Zang, who was the research director of the Baylor College of Medicine Multiple Sclerosis Comprehensive Care Center in Houston, Texas, for almost 10 years. Asians seem to have a lower incidence of the disease and there is at least one clue as to why — susceptibility genes in the major histocompatibility complex sometimes differ in patients grouped by race. "We will look at the genetics very closely," he says.

Business base

Eli Lilly is also pouring millions of dollars into China — US\$100 million over the next 5 years — but it couldn't be spending it more differently. Its rented office in Zhangjiang Hi-Tech Park employs ten people, and none does scientific research. Instead, the experienced research scientists effectively figure out how to pay other companies to do drug discovery for them, focusing initially on diabetes and

oncology. They subcontract different stages of the drug-development process to Zhangjiang's myriad CROs, in which some 300 staff members are dedicated to Lilly's work. The company's China unit, which officially opened its doors on 15 October, is calling itself a "fully integrated pharmaceutical network".

"In a smaller organization, you know the bottom line much better and you know whether you made the right decision much more quickly," says Tony Zhang, head of Eli Lilly's drug-discovery unit. "You can consider what we are doing in China an experiment. We don't have to own everything." And it is an experiment he believes will work. "There is a large pool of scientific talent, vibrant entrepreneurs, and good infrastructure," he says.

A similar experiment is being planned by Pfizer. The firm already has a China Research and Development Center that focuses on drug development and clinical trials, but it is planning to launch a similar 'virtual R&D network' that will work through partnerships with academic institutions and CROs. Such efforts by Pfizer, Roche and other companies are like "extensions of global operations," says Wang. By keeping the in-house operation small, "it's easier to open and easier to close if things don't go well".

Being able to open, close or change direction quickly is a moneysaver. Beatrijs Van Liedekerke, an associate director at Price-WaterhouseCoopers, Beijing, in charge of the business advisory group for the biotechnology, pharmaceuticals and medical-device industries, estimates that outsourcing in Chinese pharmaceutical R&D saves some 30% in costs. "They

"There are no success stories, nobody has built an Amgen or Genentech in China."

— Mingwei Wang

Started in China

While pharmaceutical companies around the world are falling over themselves to invest in China, Chinese biotech companies are struggling to be taken seriously internationally. According to Mingwei Wang, director of China's National Center for Drug Screening, the 6,600 companies in existence earlier this decade has shrunk to fewer than 3,000 after a law was implemented in 2004 requiring that drugs be made in accordance with the "good manufacturing practices" that are standard elsewhere. The remaining firms generally make generic versions of patented drugs. Just 3% of the drugs that domestic companies get approved in China each year are 'discovered in China' — and most of these are "new formulations of traditional Chinese medicine", which aren't really new, says Wang.

Now, encouraged by government support, Chinese companies are slowly gearing up towards new drugs. But the handful of true discovery companies — usually biotech firms rather than pharmaceutical companies — still gain little traction with investors, potential partners in multinational companies or the US Food and Drug Administration (FDA).

In 2002, when Qiang Yu set up Shanghai Ambrosia Pharmaceuticals with about 20 million renminbi (US\$2.5 million) from the government, "there was no drug discovery" in China, he says. "It was a blank field." Yu says he wants the first Chinese drug approved by the FDA to be his. The most promising drugs at his company are botanicals, screened from

traditional Chinese medicine libraries, which he claims work in combination to block tumour growth. "Multinationals don't want a botanical mixture. They want a single compound," says Yu. Yu is now in the second phase of fundraising but says that he is struggling to find pharmaceutical partners or investors. "Everybody is looking for later-stage products," he says. "They are demanding a

lot of data." More data might not be enough. Zensun (Shanghai) Sci & Tech, established in 2000 with 200 million renminbi (US\$30 million) from government and private sources, has taken a heart-failure drug to phase II trials at a fraction of the price it would have cost in the United States. But because the FDA sometimes doesn't accept phase II data obtained independently in China, the firm has not been able to convince potential investors of the drug's value. "When you apply to the FDA, they'll ask you to restart," says Mingdong Zhou, Zensun's chief executive. "It's a serious problem because investors evaluate us as if we were at preclinical stage, even though we are already in phase II or phase III clinical trials." **D.C.**

outsource because they don't have the expertise, they don't want to invest, or they just have too many things going on," she says. Gerald Chan, a co-founder of the Morningside, an investment company active in China, is blunter. "Drug companies had to do something to please Wall Street," he says, "and that meant outsourcing."

Roche, which in 2004 became one of the first multinationals to establish a research unit in Shanghai, is somewhere between the Eli Lilly and GSK models. Having spent some US\$70 million to \$80 million over the past four years, it now employs around 90 people in house. But it also relies heavily on CROs and partnerships with universities that, taken together, give it a total of about 250 research staff, says Chen. "You can build a drug-design chain without leaving Zhangjiang," he says.

The role of Roche's Chinese branches is partly to adapt pre-existing drugs by verifying the appropriate dose for Chinese people — essential work if the company wants to gain regulatory approval for drugs it has developed abroad. But the company also wants to make this a two-way flow, by developing a programme to identify and evaluate new drug candidates in China as far as phase II clinical trials, and pass them to Roche's global portfolio. Roche's China branch already has two cancer drugs in the preclinical stage that are ready for human trials next year. Chen agrees that outsourcing speeds up progress. "You can do several steps in parallel, and flexible resources will reduce the waiting time when

you have numerous projects with multiple compounds," he says. "By picking different CROs to do the same services in parallel, whether it's drug metabolism and pharmacokinetic studies, drug safety or oncology animal studies, pharmaceutical companies can avoid a backlog at any given point in drug development."

Wang cautions that there are limitations on just how much can be outsourced. Screening drug candidates for biological activity, for instance, requires experience that CROs are short on and this could be a concern when it comes to generating high-quality data. Very few of Zhangjiang's company's offer these types of screen. "CROs are passive. If you say 4 weeks, they'll do it. But they won't spot early

signs of deviation. It's a recipe for disaster. To interpret the data, you must be experienced."

Van Liedekerke, who has been based in China for the past two and a half years, says that China has a lot going for it, but it won't be a "deus ex machina that solves all of pharma's problems". It might solve some of them though. She gives an off-the-cuff estimate of 3–5 years before drugs discovered in China start making their way into health systems around the world. "It's just a matter of time," she says.

Drug discovery and development is becoming more complicated both scientifically and managerially, she says, and requires people with expertise in many fields to come together.

"The people that can communicate the best will win out." And whoever wins, less money will probably have been spent trying out new models in China than elsewhere.

Chan estimates that a full-time chemist in the United States costs US\$225,000 in salary and research costs per year compared with \$70,000 in China. But the price is rising quickly owing to competition and rising prices. "Shanghai used to be the most desirable spot. Now we wouldn't dare touch it," he says.

"Prices in China are going up," Chien agrees. "It will not be a bargain for much longer."

David Cyranoski is Nature's Asia Pacific correspondent.



Eli Lilly is setting up a network of subcontractors in China.



C. GARDNER/AP



Ghosts in the machine

Electronic voting machines were supposed to vanquish unreliable counts. They did not — but **David Lindley** finds that other technologies present their own problems.

In the US mid-term election of 7 November 2006, the balloting in Sarasota County, Florida, was decidedly high-tech. Voters recorded their choices on electronic touch-screen machines that had been installed following the debacle of Florida's 2000 presidential race between Republican George W. Bush and Democrat Al Gore. Then, recounts and legal actions left the United States uncertain of its next leader for more than a month before Bush was eventually declared winner of the state — and the country — by little more than 500 votes out of almost 6 million. The devices promised to prevent a repeat of that event, memorable for its images of officials solemnly peering at hanging, pregnant and dimpled chads on punch-card ballots, trying to decide which of them to count as true votes. Surely with electronic voting, any such ambiguity would be impossible?

Evidently not. By the morning of 8 November, election officials were grappling with complaints about machine glitches, unrecorded votes and 'flipped' votes allocated to the wrong candidate. Somehow 13% of voters apparently failed to register any choice at all in the closely fought congressional race between Republican Vern Buchanan and Democrat Christine Jennings. By the time the protests and the lawsuits were over, and Buchanan was officially declared the winner by 369 votes, the conclusion was painfully clear: one set of problems had been traded for another.

The effort to reform US voting technology has not been successful. Florida, for example,

threw out its old mechanical and paper-based voting technology after 2000, and switched to electronic systems, aided in part by funds allocated by the 2002 Help America Vote Act (HAVA). Critics of the move were quick to say that electronic voting machines were vulnerable to hacking and other attacks that could allow people to change the election results. And then experiences in Sarasota and elsewhere seemed to suggest that electronic machines could produce untrustworthy counts in many other ways.

As a result, Florida, New Mexico, Iowa and several other states have now rejected electronic voting and gone back to paper ballots that are marked by hand, typically by filling in a blank oval, and scanned by machine. But the cure may be worse than the disease — especially given the long history of paper ballots being lost, stolen, faked and stuffed into ballot boxes on the sly. "It's the most insecure medium there is," says Paul Herrnsen, a political scientist at the University of Maryland in College Park.

Meanwhile, some worry that arguments over the actual and alleged flaws of electronic voting systems have overshadowed more immediate concerns about usability and reliability. During the Florida 2000 election, for example, Palm Beach County's infamous 'butterfly ballot' — a split-page design with

punch-card slots running down the middle — may have led a few thousand Gore supporters to accidentally vote for Reform candidate Pat Buchanan. Misleading ballot design, it seems, may also have caused problems in Sarasota.

Adding to the confusion is the fact that in the United States, conducting elections is the responsibility of state and local governments. The result is a patchwork in which nothing is standard or uniform, something that the HAVA legislation did little to change.

Voters heading to the polls on 4 November could face any number of glitches and anomalies, and the possibility that, once again, the outcome of a close presidential race could be shrouded in uncertainty.



"Paper is the most insecure medium there is."
— Paul Herrnsen

Universal problem

Controversies over electronic voting are not unique to the United States. The Netherlands embraced electronic voting in the 1970s, in part to deal with the complexity of the nationwide system of proportional representation by which members of parliament are elected. The Dutch system came into being before hacker culture was wide-

spread, says Doug Jones, a computer scientist at the University of Iowa in Iowa City who has observed elections in the Netherlands. But in light of recent concerns, the country

R. WILSON/UNIV. MARYLAND

has decided to return to paper balloting. Brazil, by contrast, introduced electronic voting beginning in 1996, and now has a fully electronic system. In the initial stages, as many as 7% of voters trying the new system were unable to record their choices electronically, but that figure fell to less than 0.2% by 2000, and the country remains committed to its voting technology.

Lingering memories of the 2000 election, along with the chance of a close contest on 4 November, make questions about voting technology especially urgent in the United States. The debate has been particularly contentious when it comes to direct recording electronic (DRE) devices, in which a voter's choice is translated immediately into electronic data. Sarasota's iVotronic voting machines — made by Election Systems & Software based in Omaha, Nebraska — generated no record of each vote other than the electronic data. (To satisfy a HAVA audit requirement, each one did print out a paper copy of its tally once the election was finished.) In Sarasota, therefore, there was “basically no evidence that could be examined after the fact that would explain what went wrong”, says Jones.

Concerns that DRE machines are vulnerable to undetectable tampering took off in 2003, when a group of computer scientists including Avi Rubin of Johns Hopkins University in Baltimore, Maryland, and Dan Wallach of Rice University in Houston, Texas, published a report (T. Kohno *et al.* *IEEE Symposium on Security and Privacy* 9–12 May 2004) pointing out serious security flaws in the software running the AccuVote-TS, a DRE system

made by Diebold based in North Canton, Ohio. (A copy of the software had been posted on the Internet some months earlier.) As a safeguard against malfunction or tampering, the authors recommended that DRE systems should generate voter-verified paper audit trails (VVPAT) — literally, a paper display on which a voter could check that his or her vote had been cast correctly, and which could be kept as a record of the vote should any irregularities come to light.

Such a system was put in place in the May 2006 primary elections in Ohio, with printers being added to existing DRE systems. It was not a great success. A study commissioned by Ohio's Cuyahoga County found that almost 10% of the paper records were useless or absent, because, for example, the printers jammed, ran out of paper or overprinted. Ted Selker, who until June was at the Massachusetts Institute of Technology (MIT) in Cambridge as co-director of the California Institute of Technology/MIT Voting Technology Project, says that he saw similar problems when he observed elections in Nevada in 2004, one of the first occasions when a VVPAT system was added to electronic voting.

So a DRE device will record many legitimate ballots for which no paper record, or an illegible one, exists. That creates a further problem, says Michael Shamos, a computer scientist at the Carnegie Mellon University in Pittsburgh, Pennsylvania, who is also a lawyer. Election law generally states that the paper records, not the electronic data, constitute the legal ballot.

“Software fixes can really mess things up if you’re making changes at the eleventh hour.”
— Avi Rubin



2006 primary elections in Ohio, with printers being added to existing DRE systems. It was not a great success. A study commissioned by Ohio's Cuyahoga County found that almost 10% of the paper records were useless or absent,

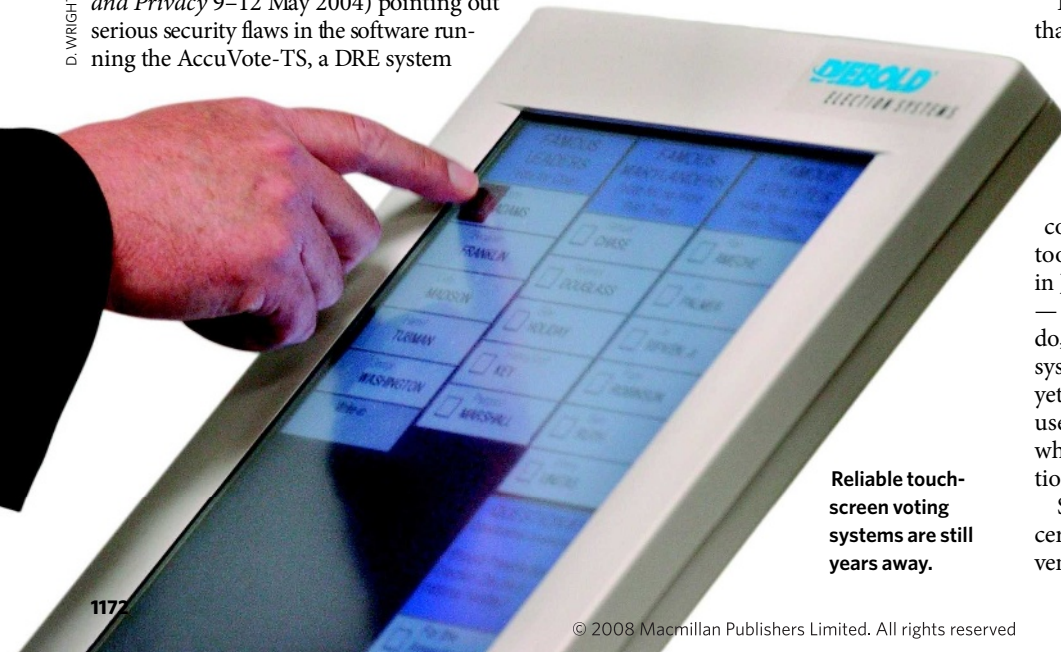
Although Rubin originally thought that adding VVPAT technology would make DRE systems workable, he now says that experience with paper-trail systems and poor implementation by the vendors has killed even that hope. And in any case, he says, VVPAT systems do nothing to solve the fundamental problem with electronic voting, which is that it depends on the reliability of software. Rubin says he is less concerned about external hacking or tampering than about software going wrong, or the risk of malicious actions from within voting-machine manufacturers. He is quick to add that he has never claimed that any manufacturer is corrupt, but thinks it's foolish to use systems that depend on them not being corrupt.

Lost in the system

Mundane programming errors may present the main danger. In the March 2008 Ohio primary election, for example, poll workers found that votes were sporadically lost during the transfer of data from individual voting-machine memory cards to the central system; after discovering the problem, election officials were able to re-read the memory cards and establish the correct totals. At the time, Premier Election Solutions — the name under which the voting machine division of Diebold now operates — claimed that the data loss resulted from a software clash with an antivirus program on the central system. On 19 August, however, Premier acknowledged in a letter to Ohio secretary of state Jennifer Brunner that it had found an internal software bug that could cause votes to be dropped when data from two memory cards were being read at the same time. Election officials have now circulated guidelines to help poll workers to circumvent the problem without altering the voting equipment.

But even if Premier devised a software patch that repaired the glitch, says Shamos, the company couldn't install it without losing certification of its system for use in federal elections. The HAVA legislation vested authority for such certification in a new Election Assistance Commission (EAC), which issued revised guidelines for the conduct of elections in December 2005 and took over the federal certification procedure in January 2007. That certification is voluntary — not every state complies — but when they do, it is not a quick process. Of the eight voting systems currently under evaluation, none has yet received EAC certification. Systems now in use were accredited under the old procedure, which was overseen by the National Association of State Election Directors.

Shamos argues for an accelerated process to certify small changes, “especially when it prevents a much more serious problem”. But Rubin



Reliable touch-screen voting systems are still years away.



Electronic voting has caused confusion for vendors, election officials and voters, but is it still more promising than paper?

J.-A. YANAK/AP

thinks that is unrealistic. He says that even large software companies, despite all the testing and checking they do, keep releasing new versions of software as bugs and problems are fixed, and then the fixes give rise to new problems that require additional fixes. A quick turnaround is “an opportunity to really mess things up if you’re making changes at the eleventh hour,” he says.

The certification process also impinges on the question of software disclosure. Both critics and advocates of DRE systems generally argue that companies should make public the software they use, which would give them greater incentive to resolve security issues. Vendors can still protect themselves by patenting their ideas and holding copyright to their code, Wallach says, and disclosure should not generate security issues for well designed software. “Only if their systems are built like garbage does disclosure become a security problem.”

But then, as Rubin points out, voting-machine design “needs to be set in concrete in order to be certified”. He imagines the case of a software security flaw or bug being discovered just before an election. “If you try to fix it, you will now be using an uncertified voting system and that would be an avenue for someone to tamper with the votes,” he says, but if you don’t fix it “you’re going to use a system that you know is bad”.

Design errors

Another concern about voting machines is that they all too often neglect basic human factors. For example, use of a printed paper record tacitly assumes that a voter confronted with an incorrect paper ballot will reject it and ask to start again. But Selker says that in practice not many voters pay attention to the paper printout, or understand its purpose.

That conclusion is supported by Sarah Everett’s PhD thesis at Rice University. Everett conducted mock elections to compare the performance of various voting technologies,

including punch cards, paper ballots and a touch-screen interface designed by Wallach and his colleagues. In some experiments, deliberate errors — omitting an election race, or flipping a voter’s selection — were introduced on the electronic system’s final review screen, which subjects were asked to check before they cast their ballot. Almost all participants said that the review screen gave them more confidence that their votes had been correctly registered. Only about one-third of them, however, actually noticed that their votes had been changed.

Of course, Jones says, if even just a few people verify their vote, such review systems can still provide warning of systematic problems, as long as election officials act. In the Sarasota case, there was “ample evidence that something was wrong before polls officially opened”. Difficulties with the machines were flagged by early voters, who under Florida law can vote ahead of election day, but their complaints brought no change in voting procedure.

The Sarasota ‘undervote’ came almost entirely from the 120,000 people who voted electronically, of whom almost 18,000 failed to register a choice in the Buchanan–Jennings congressional race. Among the roughly 20,000 absentee voters who sent in optical scan ballots, the undervote was a more typical 2.5%. Post-election testing by Florida officials and later by the Government Accountability Office

produced no conclusive evidence of machine malfunctioning, although Wallach says that the tests did not address the full range of troubles that voters reported. “We have all kinds of evidence of touch-screen malfunction and miscalibration,” he says. Touch screens must be calibrated so that they correctly connect the place where a voter touches the screen with the corresponding spot on the displayed ballot.

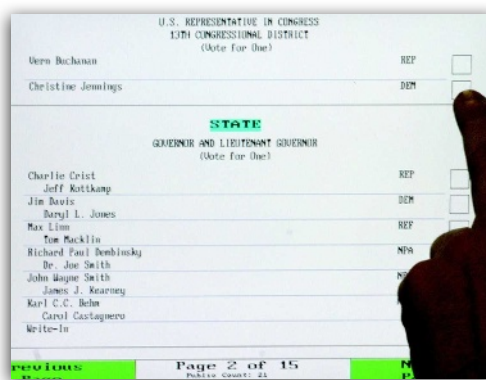
The *Sarasota Herald-Tribune*, however, suggested poor ballot design as another cause of the undervote. The Buchanan–Jennings race, in which there were no other candidates,

appeared at the top of a screen, separated from the governor/lieutenant governor race by a blue banner. The suggestion was that a voter’s eye would be drawn to the banner and the large race following it, and could easily overlook the smaller race at the top.

Selker tested this by running mock elections using the Sarasota ballot screens. He found that more

than 16% of the test subjects missed the congressional race. When he put the race at the bottom of the previous page instead, almost 19% missed it.

Selker also has an explanation for the claims of vote flipping. He cites a study by Sarah Sled of the Caltech/MIT Voting Technology Project of the 2003 election that made Arnold Schwarzenegger governor of California. Of 135 candidates, only Schwarzenegger and two others got a substantial number of votes. The six candidates immediately above or below one of



The ballot design may have led some voters in Sarasota County to miss the congressional race at the top.

G. CHAN/WFPN/UPPER/PHOTO SHOT
S. NESIUS/AP

J.-A. YANAK/AP

the three major candidates on the ballot did slightly but significantly better than the other minor candidates. Sled's conclusion was that about 0.4% of voters inadvertently voted for a candidate adjacent to their true choice.

Claims of vote flipping arise, Selker maintains, when voters mistakenly choose the wrong candidate and then, when they see a confirmation screen, assume that the machine has erred. But that's a point in favour of well-designed DRE devices, he argues: they give voters an immediate opportunity to review and correct their choices.

T. GANNAM/AP

DRE systems are also easy to use. Herrnson and his colleagues conducted experiments in which volunteers worked through a moderately complex mock election on a variety of electronic systems, and with a hand-marked paper ballot. Volunteers rated the systems on a number of criteria, including ease of use, ability to correct a mistake and confidence that their choices were accurately recorded. Two of the three touch-screen systems got most of the high marks; the third scored less well largely because it jumped to the next screen without waiting for voters to signal that they were ready to proceed.

Judged too early?

Shamos points out other advantages of DRE systems. They can more easily guide voters through complex ballots, can provide the ballot in a variety of languages, can adjust font sizes for voters with visual difficulties and are generally better for disabled voters. He says that the alarm raised over software security "was a perfectly good message, but it turned into a campaign of incredible vituperation against the concept of a computer being used in voting".

As a safeguard against the manipulation of electronic votes, Shamos says that the data representing the touch-screen image could flow directly into a non-rewritable recording medium. That record could be used to reconstruct the voter's operations on the touch screen.

Jones, however, is sceptical. Independent data capture has promise, he says, but "can you really show to me that it is independent?" What these systems aim to achieve, he says, is the electronic equivalent of a "camera over the shoulder", directly recording the voter's actions, but such techniques require data transfer and storage, and often end up being hackable themselves.

Apart from their technical challenges, proposals to improve security and provide trustworthy data back-up increase the cost of voting systems and make life more difficult for volunteer poll workers, whose training is often perfunctory. In addition, says Jones, election

officials themselves are intimidated by novel equipment, and leave instruction and trouble-shooting to representatives of the vendor. Concerns over the expense of voting equipment, along with the uncertain legislative atmosphere, combine to make "a treacherous marketplace", says Peter Lichtenheld, a spokesman for Hart InterCivic in Austin, Texas, which makes both DRE and optical-scan voting systems.

There is still no sign of resolution. Critics of DRE systems concede that no voting system is perfect — but insist that electronic voting is uniquely susceptible to failures with catastrophic, system-wide consequences. Wallach argues, for example, that defrauding a paper ballot is labour-intensive, whereas for a hacker who has the ingenuity, and the means, to mount an attack on electronic systems, "the cost to throw an election is dirt cheap". But Herrnson counters that that distinction is not as stark as it might seem. Anyone wishing to throw an election the old-fashioned way would pick on close races and would know where tampering with a few ballot boxes would swing the result.

"Some of the critics of DRE voting machines uncritically recommend paper-ballot systems without understanding that they too have shortcomings," says Jones. Even so, he thinks that optically scanned paper ballots represent the best compromise. Paper

ballots are easy to understand, he says, and although they are far from immune to fraud, "the frauds are understandable and the defences against the frauds are understandable".

For now, Rubin endorses the same system, but "not because of some love for paper systems", he says. "It's just that I think DRE systems have too many inherent problems."

Florida has now embraced optically scanned paper ballots. But on 26 August, the *Sarasota Herald-Tribune* reported that a Premier scanner being used to read absentee ballots could not upload its data to the central tallying system, and vote totals had to be entered manually. And in the District of Columbia, which uses optical scanning technology from Sequoia Voting Systems, based in San Leandro,



Electronic votes are stored on memory cards before counting.

California, early vote counts in a 9 September primary election for the city council were inflated by thousands of phantom votes that disappeared in subsequent election reports. The cause of the problem has not been definitively identified.

Moreover, Herrnson warns, the return to paper ballots inevitably raises the spectre of a repeat of Florida 2000, with a close result spawning unresolvable arguments over what does and does not constitute a legitimate mark. In place of the relentless focus on the flaws of electronic voting, he says, what is needed is a dispassionate comparison of voting systems taking security, reliability and usability into account. "The debate would benefit from a panel of election administrators and computer people who aren't bound to one position," he says. What's also needed, says Jones, is patience. In recent years, the discovery of electoral flaws has led to "irresponsible and unrealistic" expectations that new and better equipment can be put into place in a matter of months. Vendors need two years to develop voting systems to meet new standards, he says, and election officials need another two years to get the systems working reliably.

That kind of time might be available — as long as this November's election doesn't end in contention and acrimony.

David Lindley is a freelance science writer based in Alexandria, Virginia.

See Editorial, page 1149.

"Paper-ballot systems also have shortcomings."

— Doug Jones

CORRESPONDENCE

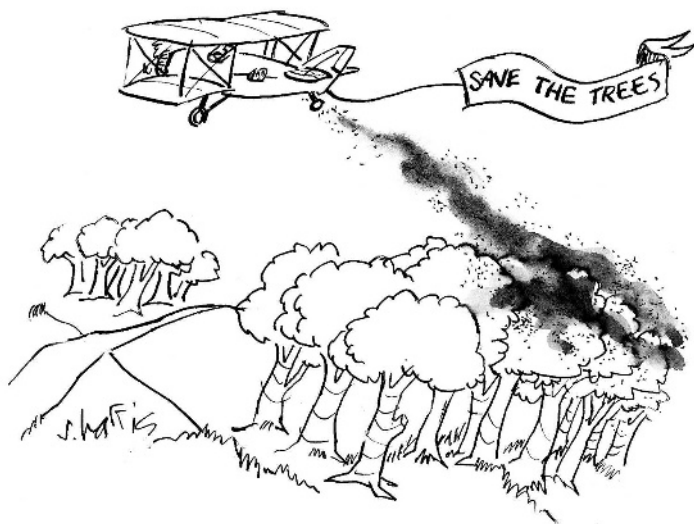
Paradox of flying to meetings to protect the environment

SIR — Scientists are becoming increasingly concerned about the environmental impact of their work. For example, in the latest of many recent reflections on the subject, bioinformatician Hervé Philippe urges us to uncouple scientific progress from economic growth (*Trends Genet.* **24**, 265–267; 2008). He adds his voice to those advising us to reduce our carbon footprint by attending fewer scientific conferences (see also the Editorial ‘Meeting expectations’ *Nature* **455**, 836; 2008).

Regular long-distance flying can easily triple an academic’s carbon footprint. During the past year, I have ‘spent’ about nine tonnes of carbon, two-thirds of this on plane trips. Yet I am a good consumer otherwise (see www.carbonfootprint.com), and I don’t even own a car. Such figures are particularly hard for field ecologists to stomach, as we hope our long-term work will highlight the environmental consequences of climate change and may ultimately influence the public and policy-makers.

Take, for instance, the dynamic field of conservation biology. Most of its best researchers are based at universities in the Northern Hemisphere, but most of their field sites are located in developing countries in the south. These hotshots and their students use up tonnes of fuel each year in commuting trips. In fact, those who are particularly renowned and most involved in environmental politics become ‘constant fliers’ who are always jetting off to field sites and meetings.

I estimate that such behaviour can potentially increase an individual’s carbon footprint to ten times their national average. Of course, plenty of businessmen have similar or larger carbon footprints, but few of them would claim that their journeys are good



for the environment. One is left wondering whether the carbon footprints of ecologists outweigh the environmental benefits of their findings and of their lobbying.

One way around this problem is to promote conservation biology on site, with improved local academic training (see, for example, <http://tinyurl.com/65kh9k>), local research leaders and fewer but longer stays by foreign visitors. With regard to environmental politics, one faces the classic dilemma between personal restraint (I stay at home and work in the garden) and energy-demanding public involvement (I fly daily to help ban overfishing).

The outcome is a personal decision that may be dictated more by ambition than by environmental awareness. Nevertheless, as a German environmental campaigner told me 15 years ago, “Industry would be all too pleased if we did not attend distant meetings because we refuse to board aeroplanes.”

David Grémillet Centre d’Ecologie Fonctionnelle et Evolutive — CNRS, 1919 route de Mende, F-34293 Montpellier, France, and FitzPatrick Institute of African Ornithology, University of Cape Town, Rondebosch 7701, South Africa
e-mail: david.gremillet@cefe.cnrs.fr

Readers are welcome to comment at <http://tinyurl.com/5hqw>.

Science students need to develop the skills of artists

SIR — In your ‘Big data’ issue, Felice Frankel and Rosalind Reid call for an “investment in visual communication training for young scientists” to prepare them for modern methods of data representation (‘Distilling meaning from data’ *Nature* **455**, 30; 2008). But the problem goes deeper than that.

Graphic artists who collaborate with scientists have often been shaped by the other of C. P. Snow’s ‘two cultures’. Although well-intentioned, many artists’ understanding of basic science is inadequate for meaningful participation in high-level technical work. Cognitive art is like commercial art and technical writing: it has never garnered respect from the artistic establishment, and its practitioners are left to fend for themselves.

From the start of schooling, distinctions are made between students with a talent for science and those with leanings towards the arts. In our technology-focused society, science receives more attention and an emphasis that does not include visual-thinking skills. Calculus, for example, is learned through symbolic operations, but portraying those procedures by using curves

and surfaces and tangents and intercepts is typically considered to be an unnecessary frill.

Thus the two cultures diverge, and if we try to reassemble them later to let one benefit the other, we have serious difficulties: the world views don’t match. Subjective ideas can be stifled by objective thought but, by the same token, physical reality can be mismanaged by well-meaning attempts at creativity.

We must indeed invest in visualization skills for science-bound students, but there should be a parallel path for science-illustrators-to-be to learn the basics of physics, chemistry and mathematics. Collaborators who understand each other’s language have a much better chance of finding the common ground they need for the cooperation they seek.

Kenneth R. Jolls Chemical and Biological Engineering, Iowa State University, 2114 Sweeney Hall, Ames, Iowa 50011-2230, USA
e-mail: jolls@iastate.edu

Military government support of science in Pakistan is illusory

SIR — Your Editorial ‘After Musharraf’ (*Nature* **454**, 1030; 2008) seems biased in its comments on the role of elected governments in Pakistan. You note that the military has ruled the country for longer than civilians have, yet you blame civilian governments for the lack of development of science.

The military governments enjoyed unaccountable and unchallenged power. For example, they were not required to reveal details of their defence budgets, and military pensions were passed over to the civilian budget when international agencies applied pressure for the military budget to be cut down (using the argument that retired military personnel are civilians).

Under General Pervez Musharraf’s rule, generals were

appointed as vice-chancellors of many of Pakistan's universities, with adverse effects on the morale of academicians and researchers. None of the military governments made a serious effort to promote science. Against this background, it is surprising to read: "Worryingly, Pakistan's governance of its science seems all set to revert to the situation that prevailed under previous elected governments."

It would be interesting to compare the rise in the military budgets under military and elected governments with the budgets for education and health. Neither military nor elected governments have ever dared to reduce the military budget in real terms.

Muhammad Naim Siddiqi, Abdul Wahab Yusufzai Department of Psychiatry, Aga Khan University, PO Box 3500, Karachi 74800, Pakistan
e-mail: naim.siddiqi@aku.edu

Doping: world agency sets standards to promote fair play

SIR — We at the World Anti-Doping Agency (WADA) find parts of Donald Berry's Commentary 'The science of doping' (*Nature* **454**, 692–693; 2008) potentially damaging to the fight against doping in sports.

WADA's accredited laboratories, including the French national laboratory LNDD (which handled the Floyd Landis case), must meet standards set by the International Standard for Laboratories (ISL) in validation methods, staff competency and chain of custody, for example. Compliance is assessed independently by bodies of the International Laboratory Accreditation Cooperation.

Berry casts doubt on the laboratory component of the anti-doping procedure as a whole, but he fails to mention that the majority of the substances reported by the anti-doping laboratories are exogenous substances not

naturally present in human beings. The development of testing procedures for endogenous substances includes samples from normal reference populations and from subjects administered with the substance under investigation, so that test-sample status and positivity criteria can be established. To determine cut-offs for the ratio of testosterone to epitestosterone (T/E), tens of thousands of athlete samples were analysed to establish reference values. To detect exogenous administration of endogenous substances (such as pharmaceutical testosterone) by isotope-ratio mass spectrometry (IRMS), validation is based on a combination of hundreds of known positive and negative samples analysed by many WADA anti-doping laboratories operating under the scrutiny of the ISL and of the International Organization for Standardization (such as ISO 17025).

Each sample test includes positive and negative quality-control samples to assess the possibility of a false result. If these samples fail, the test must be repeated. An adverse analytical finding is not reported unless the quality criteria are met.

Laboratories participate in at least four rounds of blind and one double-blind proficiency test per year; the results of each round determine the accreditation status of the laboratory. False positives mean immediate revocation of accreditation.

Berry's implication that the results of T/E ratio and IRMS analyses are interdependent is not altogether correct: according to the WADA list of prohibited substances, IRMS analysis stands alone as the basis of an exogenous testosterone finding. This is supported by legal precedents.

Contrary to Berry's suggestion that anti-doping tests may not be sufficiently specific, mass-spectrometry identification of exogenous substances relies on at least three diagnostic ions to avoid any interference

or misidentification. For immunoassays, antibodies in the initial testing and confirmation procedures must have different antigen-epitope specificity. For analytes that are too small to have two independent antigenic epitopes, two different purification methods or two different analytical methods are used.

WADA encourages its accredited laboratories to publish in peer-reviewed journals. Although complete information cannot be released without compromising the efficacy of an anti-doping test, this is made available to legal panels on request.

WADA's mandate is not to foster "a sporting culture of suspicion, secrecy and fear", as suggested by your Editorial (*Nature* **454**, 667; 2008), but to promote fair play by protecting clean athletes.

Arne Ljungqvist, Luis Horta, Gary Wadler World Anti-Doping Agency, Stock Exchange Tower, 800 Place Victoria, Suite 1700, PO Box 120, Montreal, Quebec H4Z 1B7, Canada
e-mail: violet.maziar@wada-ama.org

Doping: using flexible criteria could reduce false positives

SIR — In his Commentary 'The science of doping' (*Nature* **454**, 692–693; 2008), Donald Berry uses data from the Floyd Landis case as grounds for his claim that drug-testing practices are often based on flawed statistics. His concerns stand in contrast to the US Anti-Doping Agency's view of the outcome of the case (quoting from R. Mukhopadhyay and J. Griffiths *Anal. Chem.* **79**, 8823–8825; 2007): "It's really easy to play Monday-morning quarterback and see an i that's not dotted or a t that's not crossed, but that in no way undermined the validity or the reliability of the work that was done by the French lab."

This quote seems frivolous in the context of a class of

drug-testing practices. These are based on separation of a sample's constituents by chromatography and detection of target compounds by mass spectrometry, one of the work-horses in anti-doping research. Identification is reported as positive when the test and reference sample signals agree within a particular tolerance window.

The size of this window is not constructed with an acceptable risk of false positives in mind. Rather, fixed decision criteria hold, regardless of the quality of the laboratory or the signal properties of the target compound. However, a laboratory that produces relatively precise results should deploy stricter criteria. Likewise, target compounds should be differentiated so that information in their signals can be respected.

Always deploying the same rigid criteria leads to a probability of false positives that depends on the particular laboratory and target compound in an undefined way. This situation is frustrating because the statistical solution — flexible criteria that account for various complications — was already published and thoroughly tested five years before these rigid criteria were introduced.

So, is it ignorance of the literature or failure to understand the analytical problem at hand that underlies the ongoing usage of these arbitrary decision rules? Laboratories, as well as their clients and (re-)accrediting organizations, should start reflecting on their accountability with regard to this avoidable malpractice.

N. M. Faber Chemometry Consultancy, Goudenregenstraat 6, 6573 XN Beek-Ubbergen, The Netherlands
e-mail: nmf@chemometry.com

Contributions may be submitted to correspondence@nature.com. Published contributions are edited. Issues of interest to authors are regularly featured at *Nautilus* (<http://blogs.nature.com/nautilus>), where we welcome comments and debate.

BOOKS & ARTS

Modest heroines of time and space

The astronomical activities of two women helped to open up the heavens and ensure that a city ticked to the same time. **Patricia Fara** asks if their laborious work was valued.

Ruth Belville: The Greenwich Time Lady

by David Rooney

National Maritime Museum: 2008. 192 pp.
£12.99

The Georgian Star: How William and Caroline Herschel Revolutionized our Understanding of the Cosmos

by Michael D. Lemonick

W. W. Norton: 2008. 178 pp. \$23.95

When writers want to make science history sound exciting, they often focus on inventions, using military vocabulary to enthuse about major breakthroughs, victories over illness or unprecedented advances along the upward path of progress. Dava Sobel's *Longitude* (Walker, 1995) captivated millions of readers who enjoyed believing that a single inspired innovation — John Harrison's clock — had solved the greatest scientific problem of the age. Other historians are more circumspect. In the opinion of David Rooney, author of *Ruth Belville: The Greenwich Time Lady* and curator of timekeeping at the Royal Observatory in Greenwich, London, "New technologies don't simply replace old ones... they just add another layer of complexity to our lives." Decades after electrical time signals were introduced, he explains, many public clocks still operated mechanically, telling different times and so generating confusion rather than imparting information. How could passers-by know which of them, if any, was right?

An energetic and enterprising researcher, Rooney writes with verve, successfully infusing dramatic suspense into the everyday mundanity of horological history. He opens his saga in 1811 with a youth calling himself John Henry: his mother had fled to London from revolutionary France, but he concealed his foreign family name of Belville to secure a government job as an astronomical assistant. The book's modest heroine is John's daughter Ruth, who died in 1943 from gas poisoning after conscientiously turning down the lamp to save fuel. In between these two dates — an extraordinarily long stretch for only two generations — Rooney presents some of the fundamental stages during the agreement of universal time, spiced up with gory accounts of Jack the Ripper and a nineteenth-century terrorist whose efforts "to stop time" by blowing up a dock apparently initiated a chain of

events that influenced the twentieth-century Unabomber, Theodore Kaczynski.

The delight of Rooney's book lies not in such epoch-making events, but in his meticulous uncovering of the ordinary lives and activities that survived unchanged right through episodes of technological innovation. Before the railway system was introduced, synchronizing clocks in different parts of the country was unimportant. But after daily life had become structured by rail timetables, it became essential to make sure that people all over Britain, and eventually the entire world, were running to the same time. It was only in 1884 that international agreement was reached on Greenwich as the prime meridian, the zero line of world time. At the Royal Observatory, astronomers used

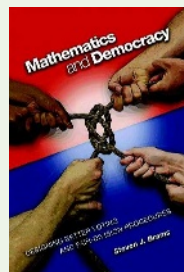
their high-precision instruments to determine the correct time. For around a century, the task of the Belville family was to ensure that this pronouncement was regularly and accurately communicated to other horological centres in London, such as watchmakers, factories and docks.

Ruth Belville spent much of her life travelling around the city, carrying in her handbag an eighteenth-century chronometer that she took to Greenwich every Monday morning so that it could be checked for accuracy against the master clock and so act as a portable standard. Well into her seventies, she trudged along the streets, a mobile purveyor of absolute time who was cheaper and more reliable than contemporary radios. She supplemented this service by

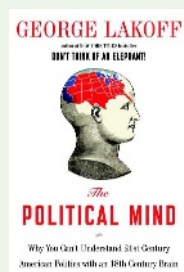


The work of Ruth Belville (above right) let London's clocks display the correct time.

FOX PHOTOS/HULTON ARCHIVE/GETTY IMAGES



The forthcoming US presidential election has spawned books about the science of voting. Political decisions can be made more democratic using social choice and game theory, allowing voters to express their preferences and enabling public goods to be divided fairly, explains Steven J. Brams in *Mathematics and Democracy: Designing Better Voting and Fair-Division Procedures* (Princeton Univ. Press, 2008). Brams has also updated his classic 1978 book, *The Presidential Election Game* (A. K. Peters, 2008), which dissects presidential campaign strategies.



Engaging emotions is key to winning votes and political arguments, argues George Lakoff in *The Political Mind: Why You Can't Understand 21st-Century American Politics with an 18th-Century Brain* (Viking, 2008). He suggests that fact-laden Democrat campaigns could benefit from using emotionally loaded phrases such as 'tax relief' or 'war on terror'.

In *Gaming the Vote: Why Elections Aren't Fair (and What We Can Do About It)* (Hill & Wang, 2008), William Poundstone seeks out the fairest voting systems and explains why the most popular candidate may not always win.

Facing decisions on topics from nuclear terrorism to climate change, the new president might consult Richard A. Muller's *Physics For Future Presidents: The Science Behind the Headlines* (W. W. Norton, 2008). Or he might turn to Joseph S. Nye's *The Powers to Lead* (Oxford Univ. Press, 2008) for tips on how to mimic the best leadership styles.

And for perking up that speech, he might consider reading *Tippecanoe and Tyler Too: Famous Slogans and Catchphrases in American History* (Chicago Univ. Press, 2008) by Jan R. Van Meter, for the stories behind 50 famous phrases, such as "nice guys finish last".

telephoning individual customers, but in 1936 the speaking clock was introduced. Known as TIM, it gave ordinary people immediate access to standardized time. In its first year, TIM took 20 million calls in London alone, from customers who dialled up to hear the recorded voice of Ethel Cain, an exchange operator picked out at an audition after her mellifluous reading of a poem by John Milton. Although there have been several other human announcers since then, this telephonic clock still speaks on, coexisting with more recent timekeepers that continuously set themselves automatically. As Rooney concludes, "New isn't necessarily better, it's just different ... stuff endures."

Rooney has the rare gift of combining the obsessiveness of an academic sleuth with the fluency of a detective novelist: not many people would attempt to write entertainingly about how the dot of the Morse code became the pip of a time signal. With its small format and subdued cover, Rooney's book superficially resembles Michael Lemonick's book *The Georgian Star*, which also uses an engaging and informative style to present a combination of history and science. Like Rooney's, Lemonick's subjects are two close relatives — in this case, the siblings William and Caroline Herschel — who also relied on an accurate scientific instrument, a telescope. There are three Georgian 'stars' in this relationship: two astronomers, here cast in heroic mode, and the planet Uranus, originally named 'George's Star' by William Herschel to gratify King George III of Great Britain and Ireland.

The Herschels worked together on the massive telescopes that helped William to become famous for his discovery of a new planet and for his work on nebulae. Disconcertingly for modern readers with egalitarian aspirations, Caroline seems to have colluded in her own marginalization, protesting that "I am nothing, I have done nothing ... a well-trained puppy-dog would have done as much".

Caroline's reputation improved towards the end of the twentieth century, when feminists commemorated her as the first woman to discover a new comet, even though this is of little scientific significance when compared with the collaboration with her brother and the star catalogue that she laboriously compiled. Caroline features prominently throughout Lemonick's book, although by the epilogue she has been reduced to a cranky old woman who helps her nephew, the Victorian astronomer John Herschel, to assume the mantle of genius passed down from his father.

Whereas Rooney presents his own original research in an accessible way, Lemonick behaves more like a journalist providing a colourful version of well-known historical and scientific material. His major source of information is the astronomy expert Michael Hoskin of the University of Cambridge, UK, and Lemonick repeatedly portrays himself as a reporter transmitting the privileged conversations he enjoyed in Hoskin's Cambridge home. Hoskin is indeed a leading authority on the Herschels, but Lemonick unspectaculously attributes to him the ability "to go beyond the straightforward facts ... and to understand their complex and remarkable personalities". At the end of *The Georgian Star*, Lemonick approvingly reproduces William Herschel's boast that, thanks to his telescope, "I have looked further into space than ever human being did before me." In contrast with this eulogy of a powerful instrument, Rooney's closing remark is that Ruth Belville "provided what no electrical wire could: the personal touch". Because she supplied what her customers wanted, her service outlasted sophisticated inventions.

Patricia Fara is the senior tutor of Clare College, University of Cambridge, Cambridge CB2 1TL, UK. Her forthcoming book is *Science: A Four Thousand Year History*.
e-mail: pf10006@cam.ac.uk

No species is an island

The Loom of Life: Unravelling Ecosystems
by Menno Schilthuizen
Springer: 2008. 220 pp. £37.50

In the summer of 1966, Harvard biologist Edward O. Wilson and his student Daniel Simberloff undertook a classic ecology experiment in Florida. They identified a number of minuscule mangrove islands, took a census of their mostly insect fauna, and then paid

an exterminator to kill all the animals on the islands with methyl bromide. Observing the islets as they became repopulated, they found that, after eight months, nearly all had regained the same number of species as they had hosted before the extermination. But most of these inhabitants were not of the same species as before.

The scientists' interventions confirmed the theory of island biogeography that Wilson and Robert MacArthur had published some years



Studies of Florida's mangrove islands have helped researchers to understand how biodiversity develops.

niches that undeniably existed". The theory puts forth a simple predictor of the abundance of various species in an ecosystem using only the rate of migration and a fundamental biodiversity number that takes into account the rate of speciation. In general, the theory shows how abundances of each species drift up and down. Sometimes new species enter the system, occasionally one will go extinct, and now and then one will speciate into two.

It is odd that the book does not give a precise definition of 'neutral' in the neutral theory, namely the assumption that any differences between the behaviour of the various species being considered — all the trees in a forest, say, or all the corals on a reef — have no overall effect in the model.

Clearly a fan of the theory, Schilthuizen makes time for its opponents, even if he undeniably calls their negative reactions "kneejerks". He sets out the evidence from both sides but avoids hashing it out, jumping on instead to ecosystem stability. All these questions are related, but the book counters expectations by becoming a survey of the great questions about the ecology of diversity instead of an argument in favour of the neutral theory.

The Loom of Life is useful. Much of the public — and even some of the professional environmental movement — knows little about the rules ecologists have posited for creating and maintaining biodiversity. They might read this slender book for a bearing on how to tackle environmental problems.

Emma Marris is a correspondent for *Nature* based in Columbia, Missouri.

earlier. It states that the number of species on an island depends on the rates of immigration and extinction, which are influenced by the distance of an island from the mainland and the island's size. Notably, the number does not depend on what the species are, nor on the roles they have in their ecosystems, known as niches.

Evolutionary biologist and science writer Menno Schilthuizen retells the story of the Wilson-Simberloff experiment at some length in *The Loom of Life*. His book is a readable, anecdotal introduction to the ecology of diversity, addressing basic questions about why there are so many different species and why some species are rare and others common.

Yet the book is hard to get into. It lacks a straightforward introduction and, with uninformative chapter titles such as 'The more the merrier', a reader who idly picks it up would need to read 50 pages to understand where all the stories are going. It is not until this point in the book — after touring through the scientific usefulness of isolated ecosystems, the history of the idea of food chains, the staggering amount of global biodiversity, and the idea that each species inhabits a slightly different niche — that the first hints of dispute enter the plot.

Schilthuizen sets out the idea that two species can coexist in the same ecosystem only if they have different niches. Evolutionary theory predicts that two species with the same living requirements and appetites would go head-to-head, and if one is even infinitesimally more efficient at exploiting their shared niche, it will, over the course of generations, outcompete and exterminate the other. Then he drops the bombshell. What about phytoplankton? Any one patch of water contains hundreds of

different species, all swimming in the same water, all competing for the same energy source: sunlight. Even with the odd difference in mineral uptake, how could so many niches exist in the uniform surface of the sea? Ecologists could go mad trying to find the minute niche differences at work; maybe there is an easier way.

After launching into the history of island biogeography, Schilthuizen introduces Stephen Hubbell's unified neutral theory of biodiversity, which Schilthuizen describes as aiming "to mimic and predict real, broad brush biodiversity patterns while ignoring the

Paris feels the polar chill

Atmosphere ... The Climate Revealed by Ice
Museum of Arts and Crafts, Paris
Until 30 April 2009

With rising levels of emotion clouding the political debate about global climate change, a new exhibition at the Museum of Arts and Crafts (*Musée des Arts et Métiers*) in Paris provides a timely reminder that accurate data are the basis of scientific objectivity. Running until April next year, *Atmosphere* shows how the past 50 years of scientific observation in the Arctic and Antarctic have provided evidence of Earth's vulnerability and changed how we view our planet. The exhibition, says geochemist Jean Jouzel, one of its two scientific commissioners, "provides a clarion call to use this carefully collected data to take informed action against climate change".

The starting point for *Atmosphere* is the International Geophysical Year of 1957–58. Objects on show in the exhibition include scientific instruments used at the poles, newspaper reports, scientific papers and books, photographs and films. Charcot, the first French polar expedition base in Antarctica, is photographed almost submerged in snow and resembles a surfacing submarine, its tricolour flag taut in the Antarctic wind. Although the focus is on raising awareness of the French contribution to polar research, the work of other nations and international collaborative groups is also presented. Polar wildlife and landscape photographs, including spectacular aurora borealis displays, document the fauna and natural beauty of the polar regions. The 130 exhibited objects range from an apparatus that simulates the aurora borealis

in the laboratory, to a 'Weasel' tracked vehicle used for transport across the snow and ice.

Jouzel's own research is included — his sampling to a depth of 3,000 metres, of polar ice that was laid down some 500,000 years ago, is crucial to our current understanding of how Earth's atmosphere has changed during this time. By analysing the composition of tiny bubbles of air trapped in polar ice, he established a correlation between temperature and the concentration of greenhouse gases, specifically carbon dioxide and methane, during successive climatic cycles. By revealing what occurred in the past, scientists can understand how best to prevent further damage to the planet. Several conferences will be held during the exhibition to alert the public to the need to change attitudes and behaviours that contribute to global warming. "But Sarah Palin is not welcome," jokes Jouzel, referring to the climate-change scepticism of the Republican vice-presidential nominee in the US election. "She is not invited."

Colin Martin is a writer based in London, UK. e-mail: cmpubrel@aol.com

See www.arts-et-metiers.net for more details.



Researchers' signs point to home from the French-Italian Concordia base in Antarctica.

G. JUGLÉ/PEV

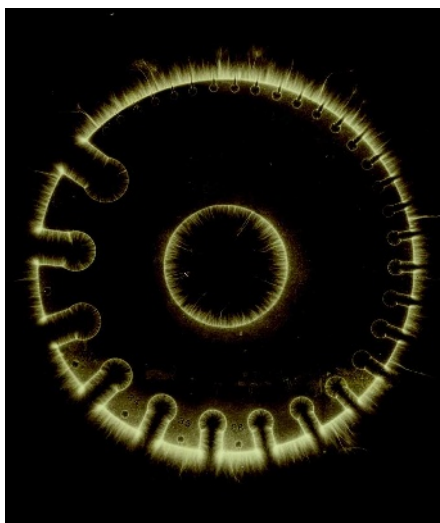
A snapshot of hidden science history

Brought to Light: Photography and the Invisible, 1840-1900

San Francisco Museum of Modern Art, California
Until 4 January 2009

The use of photography in science during the nineteenth century is traced in a new exhibition at the San Francisco Museum of Modern Art, running until 4 January 2009. *Brought to Light* showcases scientific photographs that reveal phenomena invisible to the unaided eye, from early X-rays of the body's skeleton to fields of electricity.

Scientists first considered photography as little more than a quicker way of recording the world than drawing, but soon realized that it could be used to gain access to aspects of the world that had previously been off limits. Early photographs of the Moon, the surface of the Sun, and of distant nebulae and their spectra changed astronomers' understanding of our Universe. Popular photographs of microscopic objects, such as crystals and insects, revealed new complexity in the everyday world. Indeed, in 1877, astronomer Pierre Jules César Janssen declared that the photographic plate was the "true retina of the scientist".



Hermann Schnauss's photo revealed how sparks of electricity emanate from a metal wire gauge.

Photography evolved rapidly with the new sciences. An X-ray image that Wilhelm Conrad Röntgen took of his wife's hand in 1895, for instance, was widely reproduced by others who were fascinated by the ability to see inside the human body. As exhibit curator Corey Keller writes in the catalogue,

"Today we are hardly shocked by the sight of our own bones, but we can glean from the remarkable public response just how extraordinary these pictures were when they first appeared, simultaneously suggesting an assault on privacy and conjuring the specter of death."

Keller argues that, by showing there was more going on in the world than was apparent to the unaided eye, scientific photographs inspired a fundamental shift away from faith in the human observer to faith in the authority of technology. And by the end of the nineteenth century, the heyday of the 'educated amateur' had ended. Scientists were becoming the specialized practitioners they are today, using powerful, expensive and complicated tools to explore realities that remain invisible to most.

By revealing the hidden forces at work in nature, the pictures collected in this exhibit have much in common with iconic modern images, such as the views of galaxies in the distant Universe captured by the Hubble Space Telescope. In that way, the exhibit does not just illuminate the past — it also gives a glimpse of the future.

Erika Check Hayden is a correspondent for *Nature* based in San Francisco.

ALBERTINA, VIENNA/HÖHERE GRAPHISCHE BUNDES-LEHR-UND-VERSUCHSANSTALT, VIENNA

ESSAY

Economics needs a scientific revolution

Financial engineers have put too much faith in untested axioms and faulty models, says **Jean-Philippe Bouchaud**. To prevent economic havoc, that needs to change.

Compared with physics, it seems fair to say that the quantitative success of the economic sciences has been disappointing. Rockets fly to the Moon; energy is extracted from minute changes of atomic mass. What is the flagship achievement of economics? Only its recurrent inability to predict and avert crises, including the current worldwide credit crunch.

Why is this so? Of course, to paraphrase Isaac Newton, modelling the madness of people is more difficult than modelling the motion of planets. But statistical regularities should emerge in the behaviour of large populations, just as the law of ideal gases emerges from the chaotic motion of individual molecules. To me, the crucial difference between modelling in physics and in economics lies rather in how the fields treat the relative role of concepts, equations and empirical data.

Classical economics is built on very strong assumptions that quickly become axioms: the rationality of economic agents (the premise that every economic agent, be that a person or a company, acts to maximize his profits), the 'invisible hand' (that agents, in the pursuit of their own profit, are led to do what is best for society as a whole) and market efficiency (that market prices faithfully reflect all known information about assets), for example. An economist once told me, to my bewilderment: "These concepts are so strong that they supersede any empirical observation." As economist Robert Nelson argued in his book, *Economics as Religion* (Pennsylvania State Univ. Press, 2002), the marketplace has been deified.

Physicists, on the other hand, have learned to be suspicious of axioms. If empirical observation is incompatible with a model, the model must be trashed or amended, even if it is conceptually beautiful or mathematically convenient. So many accepted ideas have been proven wrong in the history of physics that physicists have grown to be critical and queasy about their own models.

Unfortunately, such healthy scientific revolutions have not yet taken hold in economics, where ideas have solidified into dogmas. These are perpetuated through the education system: students don't question formulas they can use without thinking. Although numerous physicists have been recruited by financial

institutions over the past few decades, they seem to have forgotten the methodology of the natural sciences as they absorbed and regurgitated the existing economic lore.

The supposed omniscience and perfect efficacy of a free market stems from economic work done in the 1950s and 1960s, which with hindsight looks more like propaganda against communism than plausible science. In reality, markets are not efficient, humans tend to be over-focused in the short-term and blind in the long-term, and errors get amplified, ultimately leading to collective irrationality, panic and crashes. Free markets are wild markets.

Picture imperfect

Reliance on models based on incorrect axioms has clear and large effects. The Black-Scholes model, for example, which was invented in 1973 to price options, is still used extensively.

But it assumes that the probability of extreme price changes is negligible, when in reality, stock prices are much jerkier

than this. Twenty years ago, unwarranted use of the model spiralled into the worldwide October 1987 crash; the Dow Jones index dropped 23% in a single day, dwarfing recent market hiccups. Ironically, it was the very use of a crash-free model that helped to trigger a crash.

This time, the problem lies, in part, in the development of structured financial products that packaged subprime risk into seemingly respectable high-yield investments. The models used to price them were fundamentally flawed: they underestimated the probability that multiple borrowers would default on their loans simultaneously. These models again neglected the very possibility of a global crisis, even as they contributed to triggering one.

Surprisingly, classical economics has no framework through which to understand

'wild' markets, even though their existence is so obvious to the layman. Physics, on the other hand, has developed several models that explain how small perturbations can lead to wild effects. The theory of complexity shows that although a system may have an optimum state, it is sometimes so hard to identify that the system never settles there. This optimum state is not only elusive, it is also hyper-fragile to small changes in the environment, and therefore often

irrelevant to understanding what is going on. There are good reasons to believe that this paradigm should apply to economic systems in general and financial markets in particular. We need to break away from classical economics and develop completely different tools. Some behavioural economists and econo-physicists are attempting to do this now, in a patchy way, but their fringe endeavour is not taken seriously by mainstream economics.

While work is done to enhance models, regulation also needs to improve. Innovations in financial products should be scrutinized, crash-tested against extreme scenarios outside the realm of current models and approved by independent agencies, just as we have done with other potentially lethal industries (chemical, pharmaceutical, aerospace, nuclear energy).

Crucially, the mindset of those working in economics and financial engineering needs to change. Economics curricula need to include more natural science. The prerequisites for more stability in the long run are the development of a more pragmatic and realistic representation of what is going on in financial markets, and to focus on data, which should always supersede perfect equations and aesthetic axioms.

Jean-Philippe Bouchaud is head of research of Capital Fund Management and a physics professor at École Polytechnique in France. e-mail: jean-philippe.bouchaud@cea.fr



D. PARKINS

"Classical economics has no framework through which to understand 'wild' markets."

NEWS & VIEWS

EARTH SCIENCE

The sands of tsunami time

Stein Bondevik

The scale of the 2004 tsunami that devastated shores around the Indian Ocean has no precedent in written histories of the region. But evidence of similar events has been unearthed from the geological record.

The huge earthquake and ensuing tsunami in the Indian Ocean on 26 December 2004 killed more than 220,000 people in 11 countries. Hundreds of years of accumulated stress in the Sunda Trench was released within a few minutes and drove the Indo-Australian tectonic plate an average of 13 metres beneath the Burma–Sunda plate¹.

Written history in Indonesia — the country that suffered the most — goes back 400 years without any mention of an event of similar severity^{2,3}. So how long did it take to accumulate the energy stored along the fault that was so destructively released in 2004? In this issue, Jankaew *et al.* (page 1228)⁴ and Monecke *et al.* (page 1232)⁵ present evidence of previous tsunamis from Thailand and northern Sumatra, respectively, the results of months of hard field-work. Below sand left by the 2004 tsunami and accumulations of peat, they found sand sheets that resembled those deposited in 2004. Both studies show a sand layer dated to 600–700 years ago (ad 1300–1400) that could be the most recent full-sized predecessor of the 2004 tsunami. Such a long build-up of stress could explain the scale of that event (Fig. 1).

Ancient tsunami deposits are only rarely preserved and can be hard to find. A prerequisite is that there is some kind of a sediment trap in which the sand left by the tsunami is protected from wind, burrowing animals, running water and human activity. The authors of both papers found low areas between beach ridges, called swales, to be such sediment traps. Swales are wet and marshy, are occasionally submerged, and hold deposits of peat and other organic matter. The difference between the dark organic peat/soil and the light-coloured tsunami sand helps to trace the sand sheets. Jankaew *et al.*⁴ observed that in wet swales the 2004 tsunami sand already (as of March 2007) has a protective cover of as much as 5 centimetres of organic matter.

The buried sand sheets in the swales are similar to the sand washed up by the 2004 tsunami. But how do the authors know that the sand sheets were not deposited by other agents such as wind, rivers or storms? The reason is the geographical setting of the sites sampled by both groups^{4,5}. The sites are not exposed to



Figure 1 | The town of Lhoknga, on the northwest coast of Sumatra, before (left) and after the tsunami of 26 December 2004. These satellite photographs show the force and extent of the tsunami. The only building to survive was the mosque, seen as the prominent white structure on the right of both images. From sand layers in Thailand⁴ and northern Sumatra⁵, the best estimate is that the last tsunami of a similar scale occurred 600–700 years ago.

intense storms; rivers do not reach them; and there are no sand dunes as evidence of wind activity. Careful mapping and examination of internal structures show the sand to have been deposited from the sea side. Incorporation of plant debris and soil in the sand are evidence of strong erosion and turbulent water flow across vegetated land.

Tsunami deposits can be difficult to date, and further work needs to be done to obtain more

accurate and precise ages. Both teams used the radiocarbon method. Many of the ages they obtained are from samples of organic matter collected below the sand sheets. For these samples, there is the potential error that strong currents in the tsunami flow stripped away years of organic deposition, just before the sand settled⁶. The measured ages would then be older than the true age of the tsunami concerned. Interestingly, Jankaew *et al.*⁴ found

IKONOS IMAGES/CRISP; NATL UNIV. SINGAPORE/GOEYE

leaf fragments in the sand that were several thousand years older than the age obtained from bark fragments in the soil just below the sand layer. The age of the leaf fragments means that older material must have become mixed into the tsunami deposits, and shows the importance of selecting the right material for radiocarbon dating of such deposits.

Three older tsunami events were identified, but only the youngest of them correlates between the two regions. This sand sheet was dated to ad 1290–1400 in Sumatra⁵ and ad 1300–1450 in Thailand⁴, and could be the last big forerunner of the 2004 tsunami. The oldest sand sheet in Sumatra, dated to ad 780–900, seems to lack a corresponding sheet with the same age in Thailand. However, Jankaew *et al.*⁴ show, in one of their supplementary figures (Fig. S2c), indistinct and patchy remains of a sand sheet that they did not radiocarbon date, but that must be older than ad 1450 and younger than 395 bc. This vague sand sheet could be the same as the oldest sand sheet in Sumatra and should be subjected to dating. The third and oldest sand sheet in Thailand⁴ is about 2,200 years old, and older than the deposits studied by Monecke *et al.*⁵ in Sumatra.

If events such as the 2004 tsunami happen only every 600 years or so, there are considerable implications for urban and coastal planning in the region. The inhabitants might consider the benefits of living close to the sea as greater than the risks of a catastrophic tsunami that will not return for many generations. Also, it does not make sense to invest in and maintain a warning system for devastating tsunamis if they recur so infrequently. But smaller tsunamis may well happen more often⁷, and a warning system should save lives during such events. More geological work is also needed to measure and verify the 600-year-long recurrence intervals for exceptional tsunamis such as that of 2004.

The findings of Jankaew *et al.*⁴ and Monecke *et al.*⁵ are the first steps towards compiling a record of ancient tsunamis in the Indian Ocean, and they should inspire and encourage geologists to continue the work. There is plenty to do. To conclude that these past tsunamis were as big as the 2004 event, we need to see the same layers on other shores around the Indian Ocean. And the radiocarbon chronology of tsunamis has to be verified at different locations. ■

Stein Bondevik is in the Department of Geology, University of Tromsø, Norway, and the Faculty of Engineering and Science, Sogn og Fjordane University College, NO-6851 Sogndal, Norway. e-mail: stein.bondevik@hif.no

EVOLUTIONARY BIOLOGY

Small regulatory RNAs pitch in

Ulrich Technau

How did organismal complexity evolve at a cellular level, and how does a genome encode it? The answer might lie in differences, not in the number of genes an organism has, but rather in the regulation of gene expression.

It is commonly believed that complex organisms arose from simple ones. Yet analyses of genomes and of their transcribed genes in various organisms reveal^{1,2} that, as far as protein-coding genes are concerned, the repertoire of a sea anemone — a rather simple, evolutionarily basal animal — is almost as complex as that of a human. Grimson *et al.*³ might have the answer to this paradox. They report in this issue (page 1193) that differences in complexity between some evolutionarily ancient, morphologically simple animals and the highly sophisticated primates could be explained, at least in part, by the regulation of gene expression by small RNAs.

One of the three main classes of noncoding regulatory RNA expressed by bilaterians — organisms that have a bilateral symmetry — is the microRNAs (miRNAs). These sequences exist as RNA duplexes (roughly 21–24 nucleotides long), one strand of which usually

binds to its target messenger RNA, reducing the protein output from that message. Each miRNA has tens to hundreds of target mRNAs, adding a notable level of complexity to the regulation of gene transcription^{4,5}. In humans, hundreds of miRNAs have been identified, yet the search for related sequences in evolutionarily ancient and morphologically simple non-bilaterian animals (such as sponges, comb jellies and cnidarians) has led to only two to three putative miRNA sequences, found in the sea anemone *Nematostella vectensis*^{6–8}.

To take a closer look at the problem, Grimson *et al.*³ used deep-sequencing techniques (454 Life Sciences and Illumina sequencing machines) to analyse what was presumably the nearly full complement of small noncoding RNAs in a range of morphologically simple organisms. These included a unicellular choanoflagellate (*Monosiga brevicollis*) and several representatives of non-bilaterian

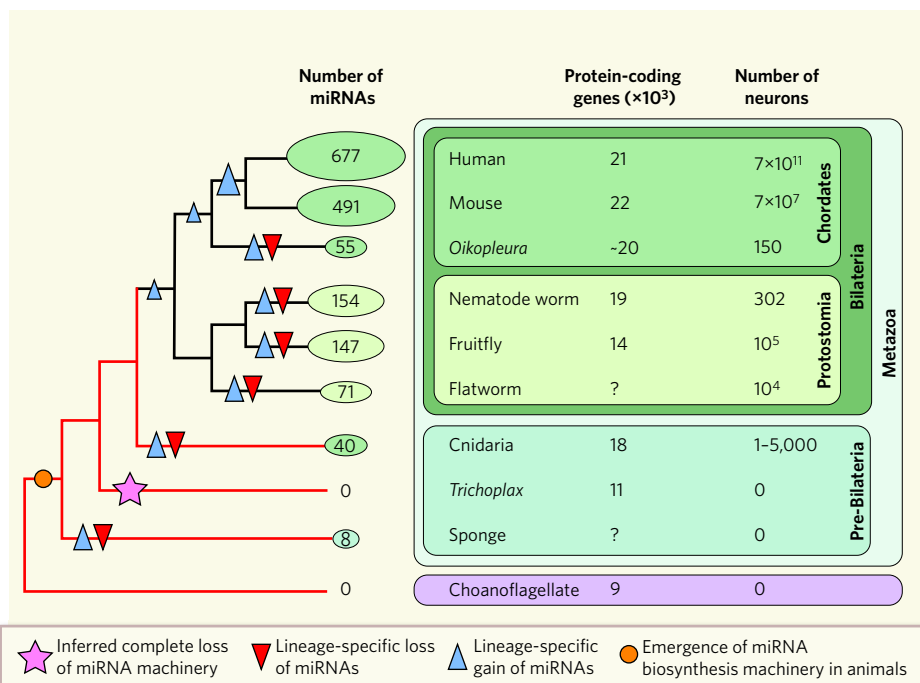


Figure 1 | Animal miRNAs and morphological complexity. Grimson *et al.*³ (data along red lines) reveal the evolutionary origin of animal miRNAs by examining organisms at the base of the animal tree. Combining their data with previous work, three different measures of complexity become apparent: the number of protein-coding genes, total number of neurons and number of miRNAs. There is relatively little correlation between morphological complexity and the number and diversity of protein-coding genes. However, miRNA number correlates well with the organism's total number of neurons. Indeed, a large proportion of vertebrate miRNAs are expressed in the nervous system. These data also show the dynamic nature of the miRNA complement in each lineage, particularly visible in rapidly evolving species (*Oikopleura* and fruitfly).

- Stein, S. & Okal, E. A. *Nature* **434**, 581–582 (2005).
- Hamzah, L., Puspito, N. T. & Imamura, F. J. *Nat. Disast. Sci.* **22**, 25–43 (2000).
- Newcomb, K. R. & McCann, W. R. *J. Geophys. Res.* **92**, 421–439 (1987).
- Jankaew, K. *et al.* *Nature* **455**, 1228–1231 (2008).
- Monecke, K. *et al.* *Nature* **455**, 1232–1234 (2008).
- Bondevik, S., Svendsen, J. I., Johnsen, G., Mangerud, J. & Kaland, P. E. *Boreas* **26**, 29–53 (1997).
- Løvholt, F. *et al.* *Nat. Hazards Earth Syst. Sci.* **6**, 979–997 (2006).

multicellular phyla: a cnidarian (*N. vectensis*), a sponge (*Amphimedon queenslandica*) and a placozoan (*Trichoplax adhaerans*). The present study³ also demonstrates the greater power of unbiased deep sequencing at near-saturation levels compared with approaches that rely on sequence similarities alone.

In the sponge, the authors find eight different miRNAs. Moreover, they report that the sea anemone has, in fact, not just three but at least 40 different miRNAs. Although, compared with the previous estimates^{6–8}, 40 miRNAs is clearly more significant, when compared with the 147 miRNAs identified in the fruitfly *Drosophila* and the 677 human miRNAs, this is not a high number. A pattern emerges, however: miRNA diversity in these three organisms seems to correlate with an increase in their relative morphological complexity (Fig. 1), even after accounting for some inflation in the number of human miRNAs owing to possible mis-annotation of certain sequences as miRNAs.

All but one of the sea anemone miRNAs seem to lack any resemblance to bilaterian miRNAs, with the one exception being almost identical in sequence to the bilaterian miRNA known as miR-100. The similarity of miR-100 sequences among sea anemone, fly and vertebrate pushes the origin of animal miRNAs back to at least the last common ancestor of sea anemones and vertebrates, some 600 million years ago. The eight sponge miRNAs do not match those of either the sea anemone or bilaterians. But, as the miRNA biosynthesis machinery is clearly fully conserved in sponges³, it seems plausible that the origin of animal miRNA dates back to the last ancestors of all animals — the Urmetazoa.

Grimson *et al.* could not convincingly detect miRNAs in either the placozoan (*Trichoplax* is a very simple organism, more like a multicellular amoeba than an animal) or the choanoflagellate, putatively the closest unicellular sister group of multicellular animals. Both of these species also seem to lack Pasha³, a protein crucial for the biosynthesis of miRNAs. Unlike other proteins of the miRNA biosynthesis machinery, Pasha's only known function is in miRNA processing. Loss of the gene encoding Pasha should therefore directly and only affect miRNA generation. Consequently, as the latest, but not uncontested, molecular phylogeny⁹ places Placozoa between the sponges and Cnidaria, the absence of miRNAs in Placozoa is probably due to secondary loss, whereby a single gene loss may have led to the complete loss of miRNAs in these organisms. Moreover, assuming a role for miRNAs in supporting higher morphological complexity, their loss, together with other gene losses⁹, may have led to the secondary morphological simplification of the placozoan body plan.

Besides miRNAs, bilaterians possess another type of (slightly longer) small RNA called piRNA. This comes in two forms, both typically expressed in the germ line — the cells

that produce gametes. Grimson *et al.* find high numbers of both piRNA subclasses in sponges and sea anemones. Strictly speaking, members of these two phyla do not have a classical germ line. Instead, multipotent stem cells, which can differentiate into diverse cell types — both somatic and germline — ensure lifelong production of gametes. Future work will undoubtedly aim to elucidate the function of piRNAs during stem-cell differentiation and gamete formation in such basal metazoans.

The finding that two basal metazoans possess relatively high numbers of miRNAs and piRNAs that have almost no related sequences in Bilateria suggests that cnidarian and sponge miRNAs have largely diversified independently of bilaterian miRNAs. But is this a rule or an exception in the evolution of these organisms? One view is that, once fixed in the population, miRNAs are seldom lost. Consequently, there would be a constant increase in diversity, correlating with an increase in morphological complexity⁷. Indeed, Grimson and colleagues' data confirm a correlation between miRNA diversity and morphological complexity as measured, for instance, by the total number of neurons in an organism (Fig. 1). But work on rapidly evolving species, such as several species of *Drosophila* and the planktonic urochordate *Oikopleura*, revealed^{10,11} both extensive gain and loss of many miRNAs over relatively short evolutionary timescales. This indicates the dynamic evolutionary nature of many miRNAs, and explains the diversity of the miRNA complement in different lineages.

What could simple organisms such as the sea anemone and sponges use their 'respectable'

number of miRNAs for? With no functional data available, one can only speculate. Many cnidarians have a rather complex, multi-stage life cycle. So, as in worms, flies and mammals, miRNAs could be involved in the timing of these different developmental stages. Another striking feature of cnidarians is their enormous capacity to regenerate. This property relies on the continuous differentiation of stem cells, possibly requiring miRNAs as part of a well-buffered molecular patterning system to ensure homeostasis. As for the role of piRNAs, in Bilateria they are usually involved in silencing mobile genetic elements called transposons. Intriguingly, compared with other animal genomes, that of the sea anemone also contains a remarkable diversity of transposable elements. Control of genome integrity by piRNAs during gamete formation might therefore be crucial for the survival of these basal metazoans over evolutionary timescales. ■

Ulrich Technau is in the Developmental Biology Section, University of Vienna, Althanstrasse 14, 1090 Vienna, Austria.

e-mail: ulrich.technau@univie.ac.at

- Putnam, N. H. *et al.* *Science* **317**, 86–94 (2007).
- Technau, U. *et al.* *Trends Genet.* **21**, 633–639 (2005).
- Grimson, A. *et al.* *Nature* **455**, 1193–1197 (2008).
- Bartel, D. P. *Cell* **116**, 281–297 (2004).
- Ambros, V. *Nature* **431**, 350–355 (2004).
- Pasquinelli, A. E. *et al.* *Nature* **408**, 86–89 (2000).
- Sempere, L. F., Cole, C. N., McPeck, M. A. & Peterson, K. J. *J. Exp. Zool. B Mol. Dev. Evol.* **306**, 575–588 (2006).
- Prochnik, S. E., Rokhsar, D. S. & Aboobaker, A. A. *Dev. Genes Evol.* **217**, 73–77 (2007).
- Srivastava, M. *et al.* *Nature* **454**, 955–960 (2008).
- Fu, X., Adamski, M. & Thompson, E. M. *Mol. Biol. Evol.* **25**, 1067–1080 (2008).
- Lu, J. *et al.* *Nature Genet.* **40**, 351–355 (2008).

ANALYTICAL CHEMISTRY

Plasma bubbles detect elements

José A. C. Broekaert

Bunsen didn't just invent the burner — he also worked on spectral analysis, in which compounds are heated in flames to produce distinctive light emissions. The latest device replaces flames with microplasmas.

Budding chemists have long enjoyed high-school experiments in which metal salts are heated in a flame to produce distinctive colours: lilac for potassium, intense yellow for sodium, brick red for calcium, and so on. But metal salts also produce distinctive colours when placed in a plasma. Staack *et al.*¹ now report in *Angewandte Chemie* that plasma can be formed at the tip of a microelectrode in solution, where it excites electrons in the solute. Upon relaxation of the electrons to the ground state, a distinctive spectrum of light is produced (the optical atomic spectrum), which can be used to detect and analyse the elements contained in the solution. This discovery could

form the basis of portable analytical devices and environmental sensors.

Optical atomic spectroscopy has an illustrious history. In the nineteenth century, Robert Bunsen and Gustav Kirchhoff nebulized solutions in flames, dispersed the emitted light with a prism, and used the intensities of lines in the resulting atomic spectra to determine the concentrations of elements in the solutes². Indeed, it was this method that led to their discoveries of caesium and rubidium. Then, in the twentieth century, many plasma radiation sources were introduced for spectrochemical analysis in the laboratory. Today, plasma atomic spectrometry is used to analyse samples from

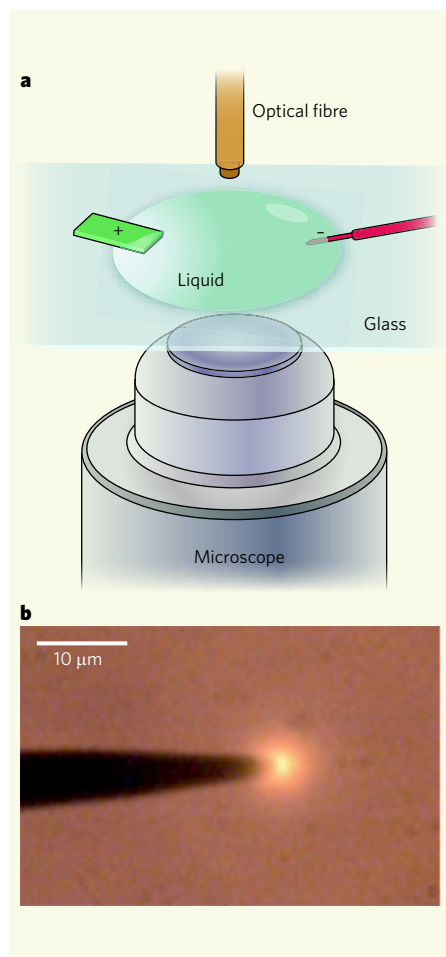


Figure 1 | Small balls of fire. Staack *et al.*¹ report a technique for detecting the elements in a solution. **a**, An electrode with a sharp tip (red) is immersed in a solution with a counter-electrode. High voltages are generated, which discharge to produce a small plasma bubble at the tip. The discharge can be observed with a microscope, and the spectrum of light produced from the plasma can be recorded using an optical fibre. The spectra produced are distinctive of the elements present in the solution. (Graphic modified from ref. 1.) **b**, This corona formed on an electrode tip immersed in blood plasma. (Image from ref. 1.)

such diverse origins as water, soil and plants, and several sampling devices allow the analysis of not only solutions, but also solid samples. The plasmas used can be generated by various methods, including electrical discharges (such as arcs and sparks) and lasers³. These methods are robust, and the resulting plasmas can be used for the accurate, precise analysis of elements in a sample, including those present in trace quantities. But there's a drawback — the instruments are bulky and expensive.

There is therefore a pressing need for smaller instruments that are cheap to make and run, and that can be easily transported for field studies. This has spawned many attempts to make minute plasma sources for use with modern, miniaturized optical spectrometers. Some success has been achieved using microwaves to induce plasmas in gas-containing channels

built into quartz wafers⁴. These inexpensive devices proved to be useful for analytical tasks — for example, they have been used successfully to determine the concentrations of toxic elements, such as mercury, in water⁵. But for this method to work, the sample has to be converted into a gas using volatilization reactions before analysis. Staack *et al.*¹ demonstrate that a minuscule plasma discharge can be produced directly in solution, and still be used to obtain atomic emission spectra from the solute.

In Staack and colleagues' device, high voltages are produced at a sharp tip (a metal electrode, a carbon wire or a quartz pipette covered with gold) when it is immersed in a solution together with a counter-electrode (Fig. 1a). The high voltage generates bubbles, formed from a combination of solute vapour, gases dissolved in the solute and solute-decomposition products. But the bubbles also take up elements contained in the solution. The bubbles are released at the tip, and go on to form a plasma discharge (Fig. 1b). The discharge measures about 10 micrometres across, and the light emitted from it contains the atomic spectral lines of the elements present in the solution. The intensities of the lines vary widely depending on such factors as the diameter and nature of the electrode; the amplitude, duration and frequencies of the voltage pulses; and the types of element present in the solution and their concentrations. Rather like a flame test, solutions of sodium chloride, tap water and blood therefore each give plasma discharges of differing colour and with different spectra.

Of course, it is still early days for this technique. The plasma discharges produced in Staack and colleagues' system have not been characterized to determine their temperatures or electron densities. Furthermore, the exact processes through which elements in the

solute are excited are not known. It will also be necessary to investigate the wavelength range over which radiation can be sampled spectrometrically — which, in turn, determines the elements that can be detected. Furthermore, the dependence of the excitation efficiency on the dimensions of the microplasmas must be quantified. Finally, all the working parameters of the technique must be precisely determined, so that any future devices can be tailored to specific analytical tasks. Nevertheless, the method has great promise: the authors show¹ that the signals produced in their experiments are proportional to the concentrations of elements in solution, and that quantities of solutes as small as a few micrograms can be detected per millilitre of solution.

The big advantage of Staack and colleagues' system¹ is likely to be its portability. The low cost of the instrument and its operation should allow it to be used in the field, and devices could even be permanently installed within the flows of the liquids to be analysed — no longer would samples have to be transported to the laboratory for analysis. Such devices therefore have potential for environmental applications, for example to detect increased concentrations of elements in rivers, or in industrial waste streams.

José A. C. Broekaert is in the Department of Chemistry, University of Hamburg, Martin-Luther-King-Platz 6, D-20146 Hamburg, Germany.
e-mail: jose.broekaert@chemie.uni-hamburg.de

1. Staack, D. *et al.* *Angew. Chem. Int. Ed.* **47**, 8020–8024 (2008).
2. Kirchhoff, G. R. & Bunsen, R. *Phil. Mag.* **20**, 89–98 (1860).
3. Broekaert, J. A. C. *Analytical Atomic Spectrometry with Flames and Plasmas* (Wiley-VCH, 2005).
4. Franzke, J. & Miclea, M. *Appl. Spectrosc.* **60**, 80A–90A (2006).
5. Engel, U., Bilgic, A. M., Haase, O., Voges, E. & Broekaert, J. A. C. *Anal. Chem.* **72**, 193–197 (2000).

CELL BIOLOGY

A *Listeria* escape trick

Grace Y. Lam and John H. Brumell

Pathogens have many ways of subverting their hosts' molecular machinery. A striking example of such a ploy comes to light from investigations of the species of bacterium that causes listeriosis.

A recent outbreak of *Listeria monocytogenes* infection in Canada¹ has to date claimed 20 lives out of 53 cases and is a stark reminder that these bacteria pose a serious threat to public health. Ingestion of contaminated food can have especially serious consequences for those at high risk — pregnant women, newborns, the elderly, and individuals with compromised immune systems.

A better understanding of *L. monocytogenes* will depend on insight into the interaction

between the bacterium and its host at the cellular level. Typically, bacteria are engulfed by macrophages, versatile agents of the immune system. They are then contained in a cellular compartment called a phagosome, which gradually becomes acidified and fuses with a compartment containing digestive enzymes, the lysosome, destroying the invader in the resulting phagolysosome. However, *L. monocytogenes* can avoid destruction by escaping through the phagosomal membrane into the

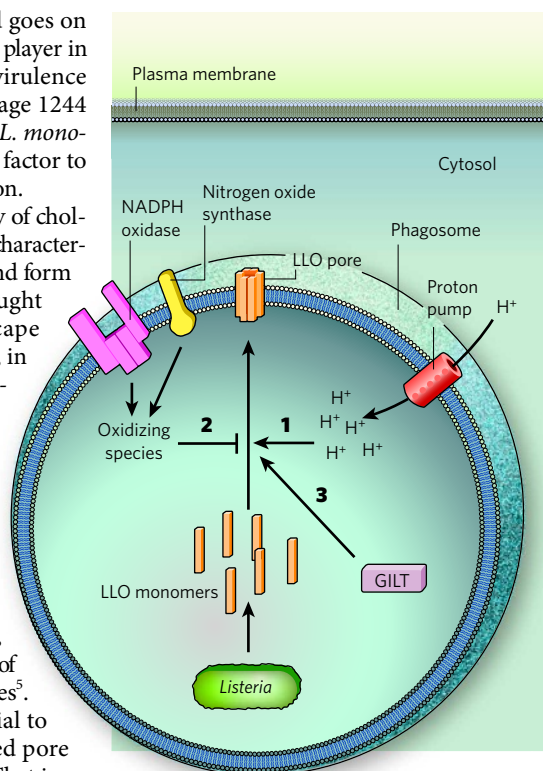
cell cytosol, where it replicates and goes on to infect neighbouring cells. The key player in the escape process is the bacterial virulence factor, listeriolysin O (LLO). On page 1244 of this issue, Singh *et al.*² reveal that *L. monocytogenes* takes advantage of a host factor to promote LLO activity during infection.

Listeriolysin O belongs to a family of cholesterol-dependent cytolysins that are characterized by their ability to oligomerize and form pores in host membranes. It is thought to mediate *L. monocytogenes* escape from phagosomes in two steps. First, in response to acidification of the phagosome, LLO undergoes structural reconfigurations that render the protein stable and active³. It then forms small pores in the phagosome, blocking acidification of this compartment and thereby preventing its fusion with lysosomes⁴. This provides a 'window of opportunity' for the second step of bacterial escape, which involves the combined actions of LLO and two phospholipase enzymes⁵. However, LLO also has the potential to kill the host cell through unchecked pore formation in the cell's membrane. That is an undesired outcome for the bacterium, which depends on the cell for its continued existence. Thus, *L. monocytogenes* has evolved many fail-safe mechanisms to limit LLO activity in the cytosol (see ref. 3 for a review).

The activity of LLO is also modulated by the phagosomal environment (Fig. 1). NADPH oxidase and nitrogen oxide synthase, two phagosomal enzyme complexes that produce antimicrobial agents — reactive oxygen and nitrogen species — inhibit escape of *L. monocytogenes*⁶. This suggests that the redox (reduction–oxidation) environment in the phagosome can control LLO activity. A single, sulphur-containing cysteine amino-acid residue in the protein may be the target of oxidation. Indeed, it has been known for some time that reducing agents are needed for optimal LLO activity *in vitro*^{7,8}.

How, then, do the bacteria overcome LLO oxidation to escape into the cytosol? Singh *et al.*² suggest that the answer lies in a host protein, γ -interferon-inducible lysosomal thiol reductase (GILT). This enzyme normally ensures the proper degradation of cysteine-containing peptides as part of another macrophage function — presentation of foreign antigen at the macrophage surface to stimulate antibody production. It seems, however, that *L. monocytogenes* takes advantage of GILT to promote LLO activity in the phagosome.

Singh *et al.* first observed that mice lacking GILT are partially protected from *L. monocytogenes* infection, in that they exhibit reduced bacterial replication in the spleen and liver. Bacterial escape from the phagosome was impaired in the macrophages of these mice, a phenomenon that was not due to differing



rates of invasion or phagosome acidification. When GILT was expressed in GILT-deficient cells, bacteria could once again escape efficiently. Furthermore, the *in vitro* introduction of mutant GILT lacking thiol reductase failed to restore the defect in bacterial escape. Conversely, using an LLO mutant in which the protein's single cysteine was replaced with alanine and which remains reduced in the phagosome, the authors discovered that this mutant confers the same bacterial escape ability in GILT-deficient macrophages as in normal cells. Together, these observations show that *L. monocytogenes* escape is impaired if LLO is not reduced by GILT.

Next, Singh *et al.* examined the mechanism of GILT action. They found that LLO reduction by GILT involves sulphur groups, as the addition of a thiol reactive reagent inhibited LLO activity in the presence of GILT. Further, they showed that the GILT–LLO interaction involves the classic two-step thiol reductase mechanism, by engineering a GILT mutant that interrupts the mechanism and results in an accumulation of the intermediate enzyme–substrate complex. When *L. monocytogenes* invaded macrophages bearing this GILT mutant, LLO bound to the mutant GILT. The data are consistent with the notion that GILT is delivered to phagosomes, where it can reduce LLO and promote its activity, allowing bacterial escape. However, because GILT was not observed in phagosomes containing the bacteria, it remains possible that GILT affects LLO (or other host or bacterial factors) elsewhere in the cell to allow bacterial escape.

It was already known that host factors affect the process by which *L. monocytogenes* escapes

Figure 1 | Regulation of the *Listeria monocytogenes* virulence factor listeriolysin O (LLO) in the phagosome of a macrophage. 1, In the normal sequence, the phagosome gradually becomes acidified through the action of proton pumps and then fuses with the lysosome (not shown), leading to bacterial degradation. But this pathway is blocked by *L. monocytogenes* through the function of LLO. As the pH drops, LLO undergoes structural reconfiguration, rendering it stable and active. It then creates pores in the phagosomal membrane, blocking acidification and thereby preventing lysosome fusion. 2, Reactive oxygen and nitrogen species are normally produced in the phagosome by NADPH oxidase and nitrogen oxide synthase enzymes, respectively. This creates an oxidizing environment in the phagosome that inhibits LLO function. 3, Singh *et al.*² propose that *L. monocytogenes* uses host γ -interferon-inducible lysosomal thiol reductase (GILT) to reduce LLO, promoting LLO's ability to form pores in the phagosomal membrane and allow subsequent bacterial escape into the cytosol.

from phagosomes⁹. Singh and colleagues' contributions² are to reveal that GILT regulates LLO activity, and to highlight the importance of the redox environment in bacterial escape. Why the bacteria require multiple levels of regulation on LLO is not clear. It is possible that exquisite control over the time and place that LLO is active is crucial in allowing bacterial escape from a specific stage in the phagolysosomal pathway. Another question raised by this work is whether other cytosol-adapted pathogens, such as *Shigella flexneri* and *Rickettsia rickettsii*, take advantage of host GILT in the same manner as *L. monocytogenes*. If so, GILT may prove to be an attractive target for therapeutic intervention.

Grace Y. Lam and John H. Brumell are at the Institute of Medical Science, University of Toronto, and the Cell Biology Program, Hospital for Sick Children, Toronto, Ontario M5G 1X8, Canada. John H. Brumell is also in the Department of Molecular Genetics, University of Toronto.

e-mails: g.lam@utoronto.ca;
john.brumell@sickkids.ca

- Public Health Agency of Canada. www.phac-aspc.gc.ca/alert-alerte/listeria/listeria_2008-eng.php
- Singh, R., Jamieson, A. & Cresswell, P. *Nature* **455**, 1244–1247 (2008).
- Schnupf, P. & Portnoy, D. A. *Microbes Infect.* **9**, 1176–1187 (2007).
- Shaughnessy, L. M., Hoppe, A. D., Christensen, K. A. & Swanson, J. A. *Cell Microbiol.* **8**, 781–792 (2006).
- Shaughnessy, L. M. & Swanson, J. A. *Front. Biosci.* **12**, 2683–2692 (2007).
- Myers, J. T., Tsang, A. W. & Swanson, J. A. *J. Immunol.* **171**, 5447–5453 (2003).
- Geoffroy, C., Gaillard, J. L., Alouf, J. E. & Berche, P. *Infect. Immun.* **55**, 1641–1646 (1987).
- Portnoy, D. A., Chakraborty, T., Goebel, W. & Cossart, P. *Infect. Immun.* **60**, 1263–1267 (1992).
- Agaisse, H. *et al. Science* **309**, 1248–1251 (2005).

CONDENSED-MATTER PHYSICS

Borderline metals

Christian Pfleiderer

The standard model of metals is found to fail in several cases. The long-sought-after marginal state in which such a breakdown occurs has been identified in a metal on the border of ferromagnetism.

In recent years a large number of metals have been discovered with properties that seem to be inconsistent with the standard model of metals — the Fermi-liquid theory. For lack of a better understanding, these materials have become known as non-Fermi-liquid metals. But the difference between a Fermi liquid and a non-Fermi liquid is like the difference between a banana and a non-banana. On page 1220 of this issue, Smith *et al.*¹ report the temperature dependence of the electrical resistivity of a non-Fermi-liquid state in the weakly ferromagnetic metal ZrZn₂, as a signature of the long-sought-after marginal Fermi-liquid state — a prediction that historically inspired the search for quantum criticality, a breakdown of Fermi-liquid theory and magnetically mediated superconductivity.

To appreciate the results of Smith *et al.*, it is helpful to recall what makes Fermi-liquid theory so special. The electrical conductivity in metals is due to the motion of the conduction electrons in response to an applied electric field. But for the charge-carrier density of around 10^{28} per cubic metre that exists in typical metals, the bare Coulomb repulsion between the conduction electrons exceeds by many orders of magnitude the strength of the electric field that causes the electric current. In other words, the movements of the conduction electrons in a weak electric field are anything but independent of one another. Fermi-liquid theory assumes that, despite these very strong interactions, it is still possible to describe the many-body excitations of the system of conduction electrons as particle-like entities (quasiparticles) that are akin to conduction electrons 'dressed' by clouds of surrounding electrons (Fig. 1).

In addition to its application to conduction electrons in metals, Fermi-liquid theory also applies to all sorts of interacting many-body systems of particles known as fermions; only one such particle can occupy a quantum state at a given time. Examples include helium-3, ³He, for which it was proposed originally, and nuclear matter in atoms and neutron stars. This makes Fermi-liquid theory and its limitations an area of wide-ranging interest.

Fermi liquids arise in systems in which particle-like excitations exist that inherit their statistical properties from the quantum-mechanical characteristics of their microscopic constituents. In metals, these constituents are the conduction electrons, which are fermions. Yet the description of the conduction electrons in terms of quasiparticles — a key property of the Fermi-liquid state — is possible only if the quasiparticles live long enough to manifest their particle-like 'personality'. In pure

compounds this implies that the lifetime of a quasiparticle is ultimately limited by scattering with other quasiparticles. The lower the temperature of the system, the fewer quasiparticles are excited, the lower the scattering rate, and the longer they live. At the limit at which the temperature falls towards absolute zero, their lifetime becomes infinite.

The Fermi-liquid state breaks down if the scattering rate between the quasiparticles grows without bound as the temperature decreases. In the simplest hypothesis, two sources could give rise to such a singular scattering rate: transverse electric currents or spin–spin interactions, neither of which can be screened by the conduction electrons. The former are important only in the microkelvin range², but spin–spin interactions are important even at high temperatures.

The marginal state identified by Smith *et al.*¹ is the result of unscreened spin–spin interactions at the border of ferromagnetism. These interactions tie in to a controversy that began in the 1920s when it was first noticed that the small, ordered magnetic moment of nickel could not be attributed to atomically localized moments only. Just like nickel, many magnetic metals exhibit a peculiar mixture of localized and itinerant magnetic moments. As a major step towards resolving this problem, it was recognized in the 1960s that the dressing clouds of quasiparticles could be paramagnons: over-damped spin waves³.

But it was immediately appreciated that these dressing clouds would imply a breakdown of the Fermi-liquid model at a zero temperature ferromagnetic critical point ($T_C = 0$) — a quantum critical point^{4,5}. This possibility inspired a research programme into the nature of quasiparticle properties in strongly correlated metals⁶ and ordering phenomena driven by these interactions, such as magnetically mediated superconductivity^{7,8}. The result was a quantitative, phenomenological model of the effects of spin fluctuations, which explained many of the contradictory properties of weakly ferromagnetic metals⁹. However, spin-fluctuation theory assumes that linear-response theory can be used self-consistently to describe strong, nonlinear interactions even at a quantum critical point^{4,5}.

But perhaps most importantly, the experimental evidence for the expected breakdown of Fermi-liquid theory at a ferromagnetic quantum critical point was missing. This theoretical hypothesis has prompted several experiments¹⁰ aimed at finding the marginal state, but the interpretation of the results is rather controversial. Smith and colleagues have now identified the marginal Fermi-liquid state in a

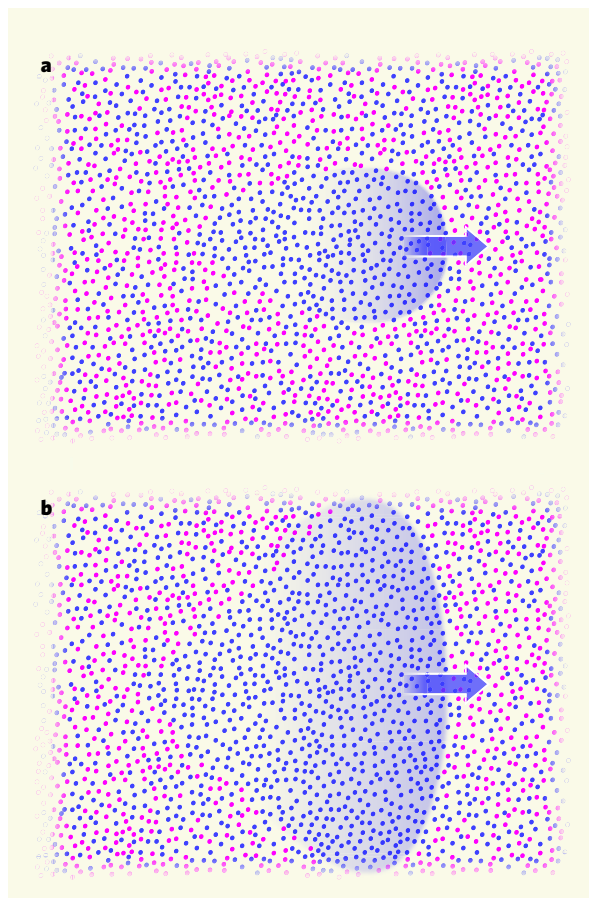


Figure 1 | Almost particles. **a**, In a normal Fermi liquid, each constituent particle (dots) is surrounded by a polarization cloud of particles (blue shading) that is dragged along by its motion (arrow). The constituent particle and 'dressing cloud', which together form what is known as a quasiparticle, act like weakly interacting particles. The blue and red dots illustrate the electrons' differing spin orientations in a nearly ferromagnetic metal: for simplicity, spin-up is red and spin-down is blue. **b**, In the marginal state, the dressing cloud depends strongly on the temperature, and is so extended that a description of the system in terms of quasiparticles is almost invalid. Smith *et al.*¹ report the identification of a marginal Fermi liquid in the weakly ferromagnetic metal ZrZn₂.

weak ferromagnet. The marginal state in these materials is reminiscent of that proposed in the context of high- T_c superconductor cuprates, but with a subtle difference: in contrast to the cuprates, the lifetime of the quasiparticles in ferromagnets involves non-local corrections.

Experimentally, the marginal Fermi-liquid state can be identified by looking at the lifetime of the quasiparticles, which is traditionally inferred from the temperature dependence of the electrical resistivity only. For a Fermi liquid, the prediction is that the electrical resistivity should scale with temperature as T^2 ; for a marginal Fermi liquid at the border of three-dimensional ferromagnetism, it should scale as $T^{5/3}$. At first sight, the variation of the electrical resistivity with temperature in the material studied by Smith *et al.*, the weakly ferromagnetic metal $ZrZn_2$, seemed to be inconsistent with Fermi-liquid theory and the marginal Fermi liquid. But the temperature dependence of the resistivity is an extremely subtle property as it also depends on the angular distribution of the scattering processes. Smith and colleagues therefore used a simple trick to determine the quasiparticle lifetime more accurately. They compared the temperature dependence of the electrical resistivity with that of the thermal

conductivity for ultra-pure specimens, and found that the quasiparticle lifetime was characteristic of the marginal Fermi-liquid state; in the marginal state, the quasiparticles live just long enough to be recognized as particles but no longer; that is, the marginal state is almost not normal (Fig. 1).

Finally, perhaps the most peculiar aspect of Smith and colleagues' work is the anomalous temperature dependence of the resistivity under pressures above p_c (the critical pressure at which ferromagnetism is entirely suppressed). The authors explain this anomalous behaviour in terms of intrinsic magnetic heterogeneities that may arise in the first-order magnetic transition at p_c (ref. 11). Yet below p_c the marginal Fermi-liquid behaviour does not seem to be affected by a small magnetic moment, which above p_c suddenly affects the resistivity strongly. This asymmetry goes along with the authors' surprising remark that the quantum critical point in $Pd_{1-x}Ni_x$ — a highly heterogeneous system that develops giant magnetic moments at the critical Ni concentration x_c (ref. 12) of a paramagnetic to ferromagnetic quantum critical point — also displays the same marginal Fermi-liquid behaviour of a pure material.

Keeping in mind that the entire analysis of the marginal state is based on the assumption that the response of the spin dynamics can be described by linear-response theory, one may speculate on whether the marginal state needs a little help to show its 'face'. This help may be a small ferromagnetic moment or a small amount of structural disorder. Without it, the normal metallic state may be much more exotic than we think. ■

Christian Pfleiderer is at the Lehrstuhl für Experimentalphysik E21, Technische Universität München, D-85748 Garching, Germany.
e-mail: christian.pfleiderer@frm2.tum.de

1. Smith, R. P. *et al.* *Nature* **455**, 1220–1223 (2008).
2. Holstein, T., Norton, R. E. & Pincus, P. *Phys. Rev. B* **8**, 2649–2656 (1973).
3. Brinkman, W. F. & Engelsberg, S. *Phys. Rev.* **169**, 417–431 (1968).
4. Hertz, J. A. *Phys. Rev. B* **14**, 1165–1184 (1976).
5. Millis, A. J. *Phys. Rev. B* **48**, 7183–7196 (1993).
6. Lonzarich, G. G. *J. Magn. Magn. Mater.* **76–77**, 1–10 (1988).
7. Mathur, N. D. *et al.* *Nature* **394**, 39–43 (1998).
8. Saxena, S. S. *et al.* *Nature* **406**, 587–592 (2000).
9. Lonzarich, G. G. & Taillefer, L. *J. Phys. C* **18**, 4339–4371 (1985).
10. Stewart, G. R. *Rev. Mod. Phys.* **73**, 797–855 (2001).
11. Uhlarz, M., Pfleiderer, C. & Hayden, S. M. *Phys. Rev. Lett.* **93**, 256404 (2004).
12. Loram, J. W. & Mirza, K. A. *J. Phys. F* **15**, 2213–2228 (1985).

STRUCTURAL BIOLOGY

Serpins' mystery solved

James C. Whisstock and Stephen P. Bottomley

Polymers of misfolded proteins underlie many diseases, including major neurodegenerative disorders. Structural data on how such aggregates of serpin proteins form answer several outstanding questions.

Ever since Alois Alzheimer and his colleagues first described abnormal aggregates of the amyloid protein in the brain of a patient who died of a form of dementia now known as Alzheimer's disease, biochemists have been fascinated by the molecular events associated with such protein misfolding. They hope that understanding the molecular basis of protein misfolding will aid the development of drugs against not only Alzheimer's disease, but also a range of disorders caused by misfolded proteins. These are known collectively as conformational diseases and include prion diseases, Parkinson's disease, Huntington's disease and several disorders known as serpinopathies¹. Whereas Alzheimer's disease and certain related disorders are caused by misfolded amyloid, serpinopathies — which include the respiratory disease emphysema, early-onset dementia and liver cirrhosis — are caused by polymerization of serpin proteins. On page 1255 of this issue, Yamasaki *et al.*² show how serpins misfold into serpinopathy-causing polymers.

The serpin story started to take shape 25 years

ago with detailed molecular and structural studies of three serpins³: α_1 -antitrypsin, a protein that reduces the inflammatory activity of the protease enzyme elastase; antithrombin, a protease inhibitor that prevents blood coagulation; and ovalbumin, from the chicken egg white. More than 85 serpin structures are now available and, cumulatively, these indicate that inhibitory serpins undergo a dramatic conformational change on contact with target proteases⁴. Notably, protease binding results in the cleavage of an exposed amino-acid sequence in the serpin molecule called the reactive-centre loop (RCL), allowing it to become an additional β -strand (s4A) in the central β -sheet — the A-sheet — of the same serpin. Certain serpins (for example, plasminogen activator inhibitor-1) can spontaneously undergo a similar conformational rearrangement without RCL cleavage to form an inactive latent protein. Conformational change in serpins has been likened⁵ to a molecular mousetrap that is sprung by a target protease and is accompanied by a large increase in protein stability.

In the mid-1980s, a sin gle-nucleotide

mutation in the gene encoding α_1 -antitrypsin (the Z-antitrypsin mutation) was discovered; the mutation results in the accumulation of highly stable, inactive polymers of this serpin in the endoplasmic reticulum of liver cells. Individuals with two copies of this mutated gene develop emphysema in early adulthood, and some develop liver cirrhosis in childhood⁶. It was suggested that the mutation might subvert the conformational flexibility required for the inhibitory activity of serpins, and that polymerization occurred through the insertion of the RCL of one serpin molecule into the A-sheet of another, and so on. This 'A-sheet polymerization' model⁶ has remained an attractive explanation for the mechanism of serpinopathies. However, it fails to readily explain some essential details, such as why serpin polymers are highly stable, and why the mutations that do not affect the stability of the normally folded protein still polymerize during folding. To address these questions, solving the structure of the disease-linked serpin polymers has remained a challenging goal, which Yamasaki *et al.*² now meet.

The structure they present is of an antithrombin dimer, from which one can easily deduce how any serpin adopting a similar configuration could propagate into a long-chain polymer. This structure reveals a remarkable twist in the tale. Instead of simple single- β -strand linkages through RCL alone, serpins can undergo a structural change known as domain swapping, whereby a main structural feature — in this case, a third of the A-sheet — forms an equivalent region in another

OBITUARY

Brian Pippard (1920–2008)

Low-temperature physicist who excelled in subtle intuitive concepts.

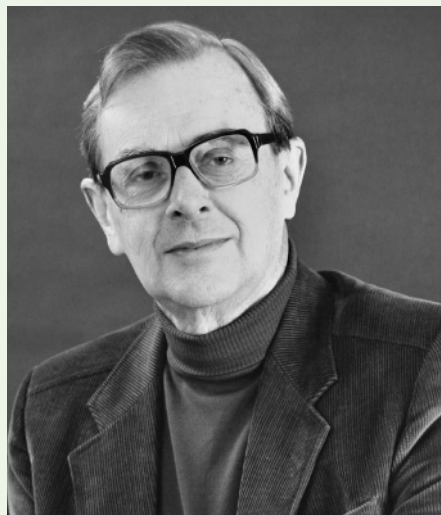
Brian Pippard, who died on 21 September, had many achievements, but will be remembered particularly as the first experimenter to map the Fermi surface of a metal, for his non-local theories of the response of normal metals and superconductors to electric fields, and as dynamic head of physics at the University of Cambridge from 1971 to 1982.

Pippard grew up in Bristol, UK, and at school proved exceptional in music as well as science. He entered Clare College, Cambridge, in 1938, and, doubting his mathematical ability, chose first to study chemistry, before switching to physics to make himself useful during the Second World War. He worked at the Great Malvern research establishment on a radar device for tracking mortar projectiles, returning to Cambridge for his PhD in 1945.

He was already familiar with the skin effect, the restriction of microwave fields to a thin surface layer in metals, and his supervisor David Shoenberg encouraged him to study it in superconductors — materials that mysteriously lose all electrical resistance below a sharply defined transition temperature. In 1935, Fritz and Heinz London had hinted that a superconductor might be a quantum fluid, a system of many electrons governed by a wavefunction similar to that of a single electron, as though they were all in the same quantum state. In such a system, the inertia of the superelectrons should lead to a purely inductive conductivity at high frequencies. By measuring the frequency and bandwidth of microwave resonators made of tin and mercury, Pippard successfully observed this effect when he cooled the resonators through their superconducting phase transitions.

Pippard's experiments were models of simplicity. For instance, the frequency of his klystron microwave source had to be stable to one part in a million, yet he achieved this using no more than draught protection, a micrometer tuner with a weight and pulley to eliminate backlash, and a stack of carefully aged high-voltage batteries as power supply.

In 1940, Heinz London had noticed that the microwave surface resistance of normal metals, instead of falling steadily with falling temperature as the resistivity did, reached a limit. Pippard confirmed this 'anomalous skin effect' in his own samples and realized that it occurred when the electron free path became larger than the skin depth, and that only those electrons moving parallel to the sample surface remained effective in responding to the electric field. This idea led quickly to a



complete theory of the effect, published in 1948 by Harry Reuter and Ernst Sondheimer, and Pippard's simple ineffectiveness concept was later applied in understanding many other phenomena, such as cyclotron resonance and ultrasonic attenuation in metals.

After his doctorate, Pippard became interested in the stability of normal metal–superconductor (NS) boundaries, and introduced the idea that their energy was determined by a characteristic coherence length. This length was later identified with that of the celebrated Ginzburg–Landau theory, and proved crucial in understanding the physics of type II superconductors, used in making superconducting magnets.

Pippard's work on the anomalous skin effect encouraged him to wonder whether, just as in normal metals the current at a given point depends on the electric field at neighbouring points through which the electrons had travelled since last being scattered, the superconducting current might likewise depend on the electric field at distant points — the range being limited by an electromagnetic coherence length. In 1953 he proposed a model of this non-local effect. It proved entirely successful; and in 1957, when John Bardeen, Leon Cooper and Robert Schrieffer published their revolutionary theory of superconductivity, they were at pains to demonstrate that it confirmed Pippard's model.

In 1954, Pippard pointed out that in the anomalous skin effect, the surface impedance of a metal is related to its Fermi surface — the theoretical boundary in momentum-space within which the conduction electrons of a metal reside — and in fact provides a measure of the curvature of this surface near

the effective electrons. In 1955, during a sabbatical year at the University of Chicago, Pippard measured the surface resistances of single-crystal samples of copper in various orientations, and on his return to Cambridge used his data to map out the Fermi surface of copper, a widely recognized tour de force. He was elected to the Royal Society in 1956.

In 1962, Pippard's student Brian Josephson independently proposed a theory of quantum tunnelling between superconductors, which subsequently earned him a Nobel prize (in 1973). Pippard was not directly involved, but later enthusiastically took up the study of the physics of SNS Josephson junctions, and of the peculiar phenomena that occur when the superconducting current is converted to normal current at an SN interface. During the same period, his book *Dynamics of Conduction Electrons* established him as an expert on all aspects of Fermi-surface phenomena, especially magnetoresistance, helicon waves, and magnetic breakdown in which electrons tunnel through momentum space from one part of the surface to another.

In 1966, Pippard became the first president of Clare Hall, informal and happy home of graduates and distinguished visitors, and of his wife Charlotte and their three daughters. In 1971, he was elected to the Cavendish chair in the department of physics, but even before that had been the moving spirit behind the much needed transfer of the Cavendish Laboratory to new buildings.

As head of department from 1971, he encouraged innovative research and tried persistently — though not always successfully — to close down areas he judged to be less worthy. He was proactive in the reform of undergraduate teaching, and insisted on lecturing ability in staff appointments. He was himself always a stimulating teacher, and his perceptive *Elements of Classical Thermodynamics* and challenging *Cavendish Problems in Classical Physics* excited and tormented generations of students. He was president of the Institute of Physics during 1974–76.

Brian could be idiosyncratic. He was suspicious of mathematical formalism, and once, when invited to deliver a keynote speech to an international audience, selected as its title 'The cat and the cream', and startled his listeners by announcing the imminent demise of solid-state physics. He was capable of great kindness, but also relished being boyishly clever: his inaugural lecture as Cavendish professor was planned around an intriguing series of bench experiments whose outcomes the assembled practitioners, young and old, were invited to predict, by show of hands. We duly got most of our predictions wrong, as he intended.

John Waldram

John Waldram was formerly at the Cavendish Laboratory, Cambridge CB3 0HE, UK.
e-mail: jrw2@cam.ac.uk

The fourth circuit element

Arising from: D. B. Strukov, G. S. Snider, D. R. Stewart & R. S. Williams *Nature* **453**, 80–83 (2008)

In 1971, Chua suggested¹ that there should in principle exist a circuit element linking electrical charge, q , and magnetic flux, ϕ . Strukov *et al.*² claim recently to have found such a link. However, here I point out that Chua's suggestion was in fact preceded by experimental evidence^{3–5} for magneto-electric effects whereby magnetic and electrical signals are interconverted. When the stimulus is magnetic and the response is electrical, the linear magneto-electric coupling constant⁵, α , is typically reported as dP/dH , where P represents the electrical polarization and H represents the applied magnetic field. The coupling constant could equally be presented as $dq/d\phi$ if divided by the permeability of free space.

In making their claim, Strukov *et al.* therefore overlook the magneto-electric literature. In magneto-electric systems, the link arises explicitly and through a simple geometry, whereas Strukov *et al.* consider a system in which magnetic signals are neither applied nor measured. Moreover, some magneto-electric systems display nonlinear coupling^{6,7}, which is necessary but not sufficient for memristance^{1,2}.

Neil D. Mathur¹

¹Department of Materials Science, New Museums Site, Pembroke Street, Cambridge CB2 3QZ, UK.

e-mail: ndm12@cam.ac.uk

Received 27 May; accepted 28 August 2008.

1. Chua, L. O. Memristor - the missing circuit element. *IEEE Trans. Circuit Theory* **18**, 507–519 (1971).
2. Strukov, D. B., Snider, G. S., Stewart, D. R. & Williams, R. S. The missing memristor found. *Nature* **453**, 80–83 (2008).
3. Astrov, D. N. The magnetoelectric effect in antiferromagnetics. *Zh. Eksp. Teor. Fiz.* **38**, 984–985; *Sov. Phys. JETP* **11**, 708–709 (1960).
4. Folen, V. J., Rado, G. T. & Stalder, E. W. Anisotropy of the magnetoelectric effect in Cr_2O_3 . *Phys. Rev. Lett.* **6**, 607–608 (1961).
5. Ascher, E., Rieder, H., Schmid, H. & Stössel, H. Some properties of ferromagnetoelectric nickel-iodine boracite, $\text{Ni}_3\text{B}_7\text{O}_{13}$. *J. Appl. Phys.* **37**, 1404–1405 (1966).
6. Hou, S. L. & Bloembergen, N. Paramagnetoelectric effects in $\text{NiSO}_4 \cdot 6\text{H}_2\text{O}$. *Phys. Rev.* **138**, A1218–A1226 (1965).
7. Scott, J. F. Mechanisms of dielectric anomalies in BaMnF_4 . *Phys. Rev. B* **16**, 2329–2331 (1977).

doi:10.1038/nature07437

Early origins and evolution of microRNAs and Piwi-interacting RNAs in animals

Andrew Grimson^{1,2}, Mansi Srivastava⁴, Bryony Fahey³, Ben J. Woodcroft³, H. Rosaria Chiang^{1,2}, Nicole King⁴, Bernard M. Degnan³, Daniel S. Rokhsar^{4,5} & David P. Bartel^{1,2}

In bilaterian animals, such as humans, flies and worms, hundreds of microRNAs (miRNAs), some conserved throughout bilaterian evolution, collectively regulate a substantial fraction of the transcriptome. In addition to miRNAs, other bilaterian small RNAs, known as Piwi-interacting RNAs (piRNAs), protect the genome from transposons. Here we identify small RNAs from animal phyla that diverged before the emergence of the Bilateria. The cnidarian *Nematostella vectensis* (starlet sea anemone), a close relative to the Bilateria, possesses an extensive repertoire of miRNA genes, two classes of piRNAs and a complement of proteins specific to small-RNA biology comparable to that of humans. The poriferan *Amphimedon queenslandica* (sponge), one of the simplest animals and a distant relative of the Bilateria, also possesses miRNAs, both classes of piRNAs and a full complement of the small-RNA machinery. Animal miRNA evolution seems to have been relatively dynamic, with precursor sizes and mature miRNA sequences differing greatly between poriferans, cnidarians and bilaterians. Nonetheless, miRNAs and piRNAs have been available as classes of riboregulators to shape gene expression throughout the evolution and radiation of animal phyla.

The RNA interference (RNAi) pathway, which processes long double-stranded RNA into small interfering RNAs and uses them to mediate gene silencing, is present in diverse eukaryotes, presumably with a role in transposon silencing or viral defence since early in eukaryotic evolution¹. Building on this basal pathway, which includes the Dicer endonuclease and the argonaute (Ago) effector protein, some eukaryotic lineages have acquired additional pathways, each using unique classes of small RNAs to guide silencing. MicroRNAs, ~21–24-nucleotide RNAs that derive from distinctive hairpin precursors, pair to messenger RNAs to direct their post-transcriptional repression². More than one-third of human genes are under selective pressure to maintain pairing to miRNAs, implying that these riboregulators influence the expression of much of the transcriptome³. Piwi-interacting RNAs are longer, ~25–30 nucleotides, with incompletely characterized biogenic pathways. In mammals and flies, piRNA expression is restricted to the germ line, where they have crucial roles in transposon defence, although one class of mammalian piRNAs, highly expressed at the pachytene stage of sperm development, has unknown function^{4,5}.

The plant and algal miRNAs have gene structure, biogenesis and targeting properties distinct from those of animals^{6–8}. These differences, considered together with the absence of miRNAs in fungi and all other intervening lineages examined, have led to the conclusion that miRNAs of animals and plants had independent origins⁶. Of the many miRNAs reported in Bilateria (Fig. 1), ~30 appear to have been present in ancestral bilaterians^{9–12}; however, none have been reported in the earliest branching animal lineages, leading to the hypothesis that bilaterian complexity might, in part, be due to miRNA-mediated regulation¹¹. Likewise, piRNAs have not been reported outside Bilateria, raising the question of whether a rich small-RNA biology is characteristic of more complex animals, or whether these small RNAs might have emerged earlier in metazoan evolution.

Diverse microRNAs of the starlet sea anemone

Eumetazoa includes the Bilateria as well as the Cnidaria. Among sequenced genomes Cnidaria is represented by the starlet sea anemone, *Nematostella vectensis*¹³. To explore whether cnidarians have miRNAs, we sequenced complementary DNA libraries generated from 18–30-nucleotide RNAs isolated from *Nematostella*. High-throughput sequencing yielded 2.9 million reads perfectly matching the *Nematostella* genome (Fig. 2a). To identify miRNAs, we considered properties that have proved useful for distinguishing bilaterian miRNAs from other types of small RNAs represented in sequencing data^{14,15}. The first criterion was the presence of reads mapping to an inferred RNA hairpin with pairing characteristics of known miRNA hairpins. The second was the presence of reads from both arms of the

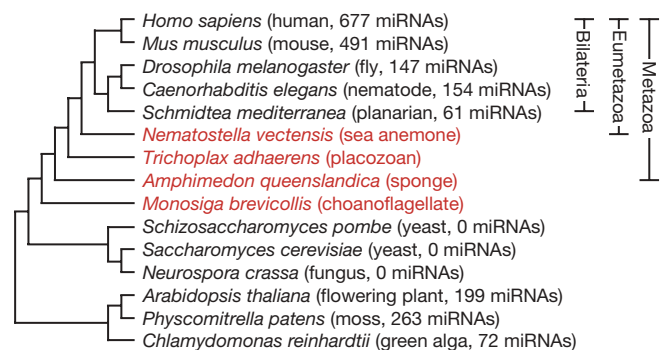


Figure 1 | Phylogenetic distribution of annotated miRNAs. Cladogram of selected eukaryotes, with organisms investigated in this study indicated in red. Branching order of Bilateria is according to ref. 28 and the references therein, and that of basal Metazoa is according to ref. 17 (Supplementary Discussion). Annotated miRNA tallies are from miRBase (v10.1)²⁹.

¹Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, Massachusetts 02142, USA. ²Howard Hughes Medical Institute, Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. ³School of Integrative Biology, University of Queensland, Brisbane 4072, Australia. ⁴Department of Molecular and Cell Biology and Center for Integrative Genomics, University of California at Berkeley, Berkeley, California 94720, USA. ⁵Department of Energy, Joint Genome Institute, Walnut Creek, California 94598, USA.

hairpin that, when paired to each other, formed a duplex with 2-nucleotide 3' overhangs. This duplex corresponds to an intermediate of miRNA biogenesis in which the miRNA and opposing segment of the hairpin, called the miRNA*, are excised from the hairpin through successive action of Drosha and Dicer RNase III endonucleases². The third criterion was homogeneity of the miRNA 5' terminus. Because pairing to miRNA nucleotides 2–8 is crucial for target recognition³, reads matching bilaterian miRNAs display less length variability at their 5' termini than at their 3' termini^{14,15}.

As exemplified by *mir-2024d* (Fig. 2b, c), 40 distinct *Nematostella* loci met these criteria (Fig. 2d and Supplementary Data 1; identical hairpins were not counted because they might have arisen from genome-assembly artefacts). Additional features, not used as selection criteria, resembled those of bilaterian miRNAs², thereby increasing confidence in our annotations. For example, the loci usually mapped between annotated protein-coding genes (31 loci) or within introns in an orientation suitable for processing from the pre-mRNA (8 loci). The *Nematostella* miRNAs also had a tight length distribution (centring on 22 nucleotides, Fig. 2d), and five groups of miRNAs (corresponding to 13 miRNAs) mapped near to each other in an orientation suitable for production from the same primary transcript (Supplementary Data 1), as occurs in bilaterians². With the exception of two miRNA pairs (*mir-2024a,b* and *mir-2024f,d*), the *Nematostella* miRNAs had unique sequences at nucleotides 2–8, suggesting notable diversity of miRNA targeting in this simple animal.

Previous studies that explored the possibility that cnidarians might have miRNAs searched for *Nematostella* homologues of the ~30 miRNA families broadly conserved within the Bilateria by probing RNA blots and examining candidate hairpin sequences^{11,12}. These studies reported the possible presence of miR-10, miR-33 and miR-100 family members in *Nematostella*. None of our reads matched the proposed miR-10, miR-33 or miR-100 homologues, and none matched the proposed hairpin precursors of miR-10 or miR-33. Such discrepancies were not unexpected, because detection of distantly related miRNAs by hybridization is prone to false-positives,

and many genomic sequences can fold into hairpins. However, one of the newly identified miRNAs arose from the hairpin of the reported miR-100 homologue. The actual miRNA was offset by one nucleotide compared to bilaterian miR-100 family members (Fig. 2e). Because miRNA-targeting is defined primarily by nucleotides 2–8, this offset is expected to alter target recognition substantially, with the *Nematostella* version primarily recognizing mRNAs containing CUACGGG and UACGGGA heptanucleotide sites and the bilaterian versions recognizing mRNAs with two different sites, UACGGGU and ACGGGUA³.

Despite this wholesale shift in their predicted targeting, the *Nematostella* and bilaterian versions of miR-100 had similarity throughout the RNA, suggesting common origins (Fig. 2e). This result confidently extended the inferred origin of metazoan miRNAs back to at least the last common ancestor of these eumetazoans. Systematic comparison to annotated miRNAs did not reveal any additional *Nematostella* miRNAs with similarity exceeding that of shuffled control sequences (Supplementary Fig. 1). Although the short length of miRNAs may cause sequence divergence to obscure common ancestry, it is noteworthy that only one of the 40 *Nematostella* miRNAs appeared homologous to extant bilaterian miRNAs, and even this one seemed to have profoundly different targeting properties.

MicroRNAs near the base of the metazoan tree

To determine whether miRNAs might be present in more deeply branching lineages, we generated 2.5 million genome-matching reads from the small RNAs of the demosponge *A. queenslandica*, a poriferan thought to represent the earliest diverging extant animal lineage^{16,17} (Figs 1 and 3a). Eight miRNA genes were identified in *Amphimedon* adult and embryo samples (Fig. 3b and Supplementary Data 2), exemplified by *mir-2018* (Fig. 3c). Six mapped between annotated protein-coding genes; two fell within introns. As is typical for bilaterian miRNAs² and is also found in *Nematostella* (Fig. 2d), reads from one arm of the hairpin usually greatly exceeded those from the other arm, enabling unambiguous

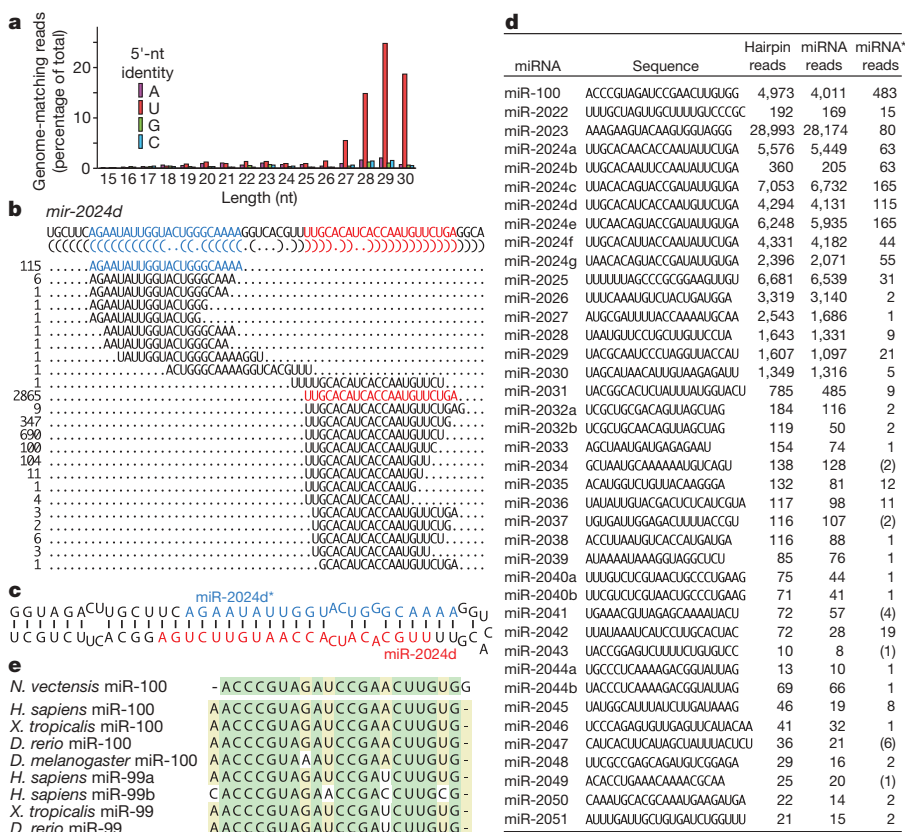


Figure 2 | The miRNAs of *N. vectensis*. **a**, Length distribution of genome-matching sequencing reads representing small RNAs, plotted by 5'-nucleotide (nt) identity. Matches to ribosomal DNA were omitted. **b**, Sequencing reads matching the *mir-2024d* hairpin. The sequence of the *mir-2024d* hairpin is depicted above the bracket-notation of its predicted secondary structure. The sequenced small RNAs mapping to the hairpin are aligned below, with the number of reads shown on the left, and the designated miRNA and miRNA* species coloured red and blue, respectively. Analogous information is provided for the other newly identified miRNAs (Supplementary Data 1). **c**, Predicted secondary structure of the *mir-2024d* hairpin, indicating the miRNA and miRNA* species. **d**, The 40 *Nematostella* miRNAs. MicroRNA read counts include those sharing the dominant 5' terminus but possessing variable 3' termini. Occasionally the only sequenced miRNA* species corresponded to a variant miRNA species rather than the major species (counts in brackets). **e**, Alignment of miR-100 homologues (*Danio rerio*, *D. rerio*; *Xenopus tropicalis*, *X. tropicalis*).

annotation of the miRNA and miRNA* (Fig. 3b). However, the number of reads from the two arms of the *mir-2015* hairpin did not differ substantially, suggesting that each might have similar propensities to enter the silencing complex and target miRNAs. Moreover, the species from the 3' arm (miR-2015-3p) dominated in adult tissue, whereas the one from the 5' arm (miR-2015-5p) dominated in embryonic tissue (Fig. 3d), supporting the notion that this single hairpin produces two distinct miRNAs, and implying an intriguing, developmentally controlled differential loading into the silencing complex.

In *Amphimedon*, pre-miRNA hairpins were larger than most of those of other metazoans (Fig. 3e). The *Nematostella* pre-miRNAs (including *mir-100*) fell at the other end of the spectrum, with a median length less than that of bilaterian pre-miRNAs (Fig. 3e). None of the *Amphimedon* miRNAs shared significant similarity with any previously described miRNAs (Supplementary Fig. 1), or with the miRNAs found in *Nematostella*. This observation, combined with their unusually large pre-miRNA hairpins, raised the possibility of an origin independent from that of eumetazoan miRNAs. Arguing against this possibility, we found *Amphimedon* homologues of Drosha and Pasha proteins (Table 1), which recognize the miRNA primary transcript and cleave it to liberate the pre-miRNA hairpin¹⁸. Homologues of these proteins appeared to be absent in all lineages outside the Metazoa, indicating a single origin for these processing factors early in metazoan evolution and implying a single origin for their miRNA substrates.

A third animal lineage branching basal to the Bilateria is Placozoa, represented by the sequenced species *Trichoplax adhaerens*¹⁷. Although earlier analyses of mitochondrial genes suggested that *Trichoplax* diverged before *Amphimedon*, genomic data indicate that *Trichoplax* had a common ancestor with cnidarians and bilaterians more recently than with *Amphimedon*¹⁷ (Fig. 1 and Supplementary Discussion). Our study of *Trichoplax* small RNAs failed to find miRNAs, despite acquiring many more reads than required to identify miRNAs in all other animals and plants examined (Supplementary Figs 2 and 3). Thus, despite the formal possibility that *Trichoplax* miRNAs are expressed at levels so low that we failed to detect them, we favour the hypothesis that all miRNA genes have been lost in this lineage. *Trichoplax* is thought to have derived from a more complex ancestor, having lost, for example, the hedgehog and Notch signalling pathways¹⁷. Supporting our hypothesis, no Pasha homologue was found in the *Trichoplax* genome, although we did find the core RNAi proteins—argonaute and Dicer—suggesting the

production and use of small interfering RNAs (Table 1). Drosha, which partners with Pasha during miRNA biogenesis¹⁸, was found also but might be required in the absence of miRNAs for ribosomal RNA maturation¹⁹. Of the proteins involved in canonical miRNA biogenesis, Pasha is the one without known functions outside the miRNA pathway, and it was the one that appeared to have been discarded, together with all miRNAs, from the *Trichoplax* genome (Table 1).

We also sequenced small RNAs from the single-celled organism *Monosiga brevicollis* (Supplementary Fig. 2), which represents the closest known outgroup to the Metazoa²⁰. We failed to detect any plausible miRNAs, a result consistent with our subsequent finding that *Monosiga* seems to lack all genes specific to small-RNA biology (Table 1). The absence of Dicer and argonaute seemed to be derived rather than ancestral, as the common ancestor of *Monosiga* and metazoans possessed these core RNAi proteins¹ (Table 1). The possibility that the absence of miRNAs in *Monosiga* might likewise be derived prevented us from setting an early bound on the origin of metazoan miRNAs.

In summary, miRNAs appear to have been available to shape gene expression since at least very early in animal evolution. Nonetheless, the numbers identified in simpler animals (8 unique miRNAs in *Amphimedon* and 40 in *Nematostella*) were lower than those reported in more complex animals (Fig. 1). Although miRNAs expressed only under specific conditions or at restricted developmental stages were possibly missed in these and other animals, our results are consistent with the idea that increased organismal complexity in Metazoa correlates with the number of miRNAs and presumably with the number of miRNA-mediated regulatory interactions.

Piwi-interacting RNAs in deeply branching animals

We next turned to the possibility that piRNAs also might have early origins. Piwi proteins, the effectors of bilaterian piRNA pathways, are found in diverse eukaryotic lineages (although not in plants or fungi, Table 1), implying their presence in early eukaryotes¹. In cases characterized, however, the small RNAs associated with non-metazoan Piwi proteins resemble siRNAs more than bilaterian piRNAs (deriving, for example, from Dicer-catalysed cleavage of long double-stranded RNA²¹), raising the question of when piRNAs of the types found in Bilateria might have emerged. The genomes of both *Amphimedon* and *Nematostella*, but not that of *Trichoplax*, encode Piwi proteins (Table 1) and express many ~27-nucleotide RNAs with a 5'-terminal uridine (5'-U) (Figs 2a and 3a)—features reminiscent of

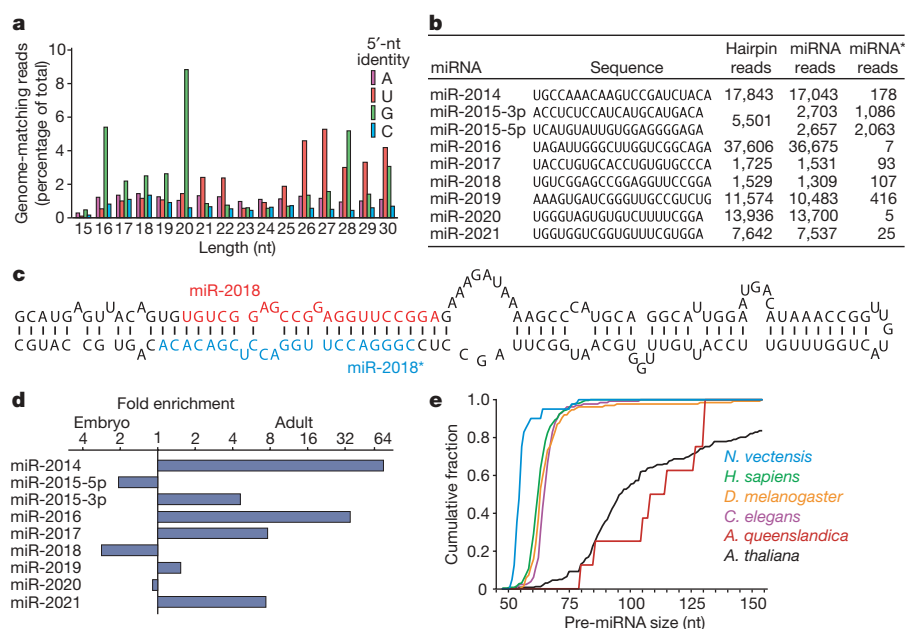


Figure 3 | The miRNAs of *Amphimedon queenslandica*. **a**, Length distribution of genome-matching sequencing reads representing small RNAs, plotted by 5'-nucleotide identity. Matches to ribosomal DNA were omitted. **b**, The *Amphimedon* miRNAs, shown as in Fig. 2d. Information analogous to that of Fig. 2b is provided for these miRNAs (Supplementary Data 2). **c**, Predicted secondary structure of the *mir-2018* hairpin. **d**, Relative expression of *Amphimedon* miRNAs, as indicated by sequencing frequency from adult and embryo samples. **e**, Cumulative distributions of pre-miRNA lengths from miRNA transcripts of the species indicated. *Amphimedon* pre-miRNAs were significantly larger than those from any other animal species examined ($P < 10^{-5}$, Wilcoxon rank-sum test), whereas those from *Nematostella* were significantly smaller ($P < 10^{-5}$).

Table 1 | The small-RNA machinery of representative eukaryotes

Species	Ago	Piwi	Dicer	Drosha	Pasha	Hen1
<i>Homo sapiens</i>	4	4	1	1	1	1
<i>Drosophila melanogaster</i>	2	3	2	1	1	1
<i>Caenorhabditis elegans</i> *	5	3	1	1	1	1
<i>Nematostella vectensis</i> †	3	3	2	1	1	1
<i>Trichoplax adhaerens</i> ‡	1	0‡	5	1	0§	0‡
<i>Amphimedon queenslandica</i> †	2	3	4	1	1	2
<i>Monosiga brevicollis</i>	0‡	0‡	0‡	0	0	0‡
<i>Saccharomyces cerevisiae</i>	0‡	0‡	0‡	0	0	0‡
<i>Schizosaccharomyces pombe</i>	1	0‡	1	0	0	0‡
<i>Arabidopsis thaliana</i>	10	0‡	4	0	0	2
<i>Physcomitrella patens</i>	6	0‡	5	0	0	1
<i>Chlamydomonas reinhardtii</i>	2	0‡	3	0	0	1

* Omitted is a nematode-specific clade of proteins related to the Ago and Piwi protein families but distinct from both²⁷.

† Protein sequences are listed in Supplementary Data 3.

‡ Inferred loss based on presence in earlier-diverging lineages.

§ Inferred loss based on presence in earlier-diverging lineages when assuming that *Amphimedon* diverged before *Trichoplax* (Supplementary Discussion).

|| Ago and Dicer, but not Piwi, Drosha, Pasha or Hen1, were also identified in each of the additional fungal species examined (*Aspergillus nidulans*, *Neurospora crassa* and *Sclerotinia sclerotiorum*).

piRNAs in vertebrates and flies⁵. Moreover, 45% of *Nematostella* 5'-U 27–30-nucleotide RNAs originated from only 89 genomic loci (together comprising 0.4% of the genome), the largest of which was 62 kilobases, and essentially all of these small RNAs derived from one strand of each locus (Fig. 4a and Supplementary Table 3). In these respects the genomic loci producing a large fraction of the *Nematostella* reads closely resembled the loci producing bilaterian piRNAs, particularly the pachytene piRNAs⁵. We observed a similar clustering of genomic matches of *Amphimedon* 5'-U 24–30-nucleotide RNAs, although the loci were smaller and accounted for fewer reads (10% of the reads originating from 73 loci comprising 0.2% of the genome, Supplementary Table 4).

Another characteristic of piRNAs is that they undergo Hen1-mediated methylation of their terminal 2' oxygen²². To test for this modification, we treated RNA from *Nematostella* and *Amphimedon* with periodate and then re-sequenced from both treated and untreated samples (Supplementary Fig. 4). Piwi-interacting RNAs and other RNAs modified at their 2' oxygen remain unchanged with this treatment and are sequenced, whereas those with an unmodified 2',3' cis-diol are oxidized, which renders them refractory to sequencing²³. In contrast to the *Amphimedon* miRNAs and many of the *Nematostella* miRNAs (Supplementary Tables 1 and 2), reads corresponding to the candidate piRNA clusters in both *Nematostella* and *Amphimedon* were not reduced after treatment (Supplementary Tables 3 and 4), indicating that their terminal 2',3' cis-diol was modified. This modification, considered together with their other features characteristic of vertebrate and fly piRNAs, including the length of 25–30 nucleotides, the 5'-U bias, and the single-stranded, clustered organization of their genomic matches, provided evidence that these small RNAs represented piRNAs of *Nematostella* and *Amphimedon*.

The piRNAs were the type of small RNAs most abundantly sequenced in *Nematostella* and *Amphimedon* (Figs 2a and 3a, and Supplementary Discussion). A similar phenomenon is observed in mammalian testes, in which the pachytene piRNAs greatly outnumber the miRNAs and initially obscured detection of a second class of mammalian piRNAs, which resemble the most abundant *Drosophila* piRNAs with respect to both their biogenesis and their apparent role in suppressing transposon activity²⁴. Most of the *Nematostella* and *Amphimedon* genomic loci with clustered piRNA matches resembled the first class of piRNAs, in that they tended to fall outside of annotated genes ($P < 10^{-3}$, Wilcoxon rank-sum test) and spawned piRNAs predominately from only one DNA strand (>99% and 96% from one strand, *Nematostella* and *Amphimedon*, respectively). To determine whether the second class of piRNAs might also exist in deeply branching lineages, we analysed the sequences from periodate-treated samples, focusing on the minority that matched annotated protein-coding

genes (Fig. 4b). As expected for class II piRNAs, these piRNAs did not have such a strong tendency to match only one strand of the DNA (62% and 64% antisense for *Nematostella* and *Amphimedon*, respectively). Moreover, among the predicted coding regions with the most matches to the piRNAs, a significant fraction (18 of 50 in *Nematostella*, $P < 10^{-3}$; 12 of 40 in *Amphimedon*, $P = 0.03$, Supplementary Tables 5 and 6) were homologous to transposases.

Having found small RNAs resembling bilaterian class II piRNAs we looked for evidence that they were generated through the same feed-forward biogenic pathway^{4,25}. In this pathway, primary piRNAs from transcripts antisense to transposable elements pair to transposon messages and direct their cleavage. This cleavage defines the 5' termini of secondary piRNAs generated from the transposon message, and these secondary piRNAs pair to piRNA transcripts, directing cleavage and thereby defining the 5' termini of additional piRNAs resembling the primary piRNAs. Because the primary piRNAs typically begin with a 5'-U and direct cleavage at the nucleotide that pairs to position 10, the secondary piRNAs typically have an A at

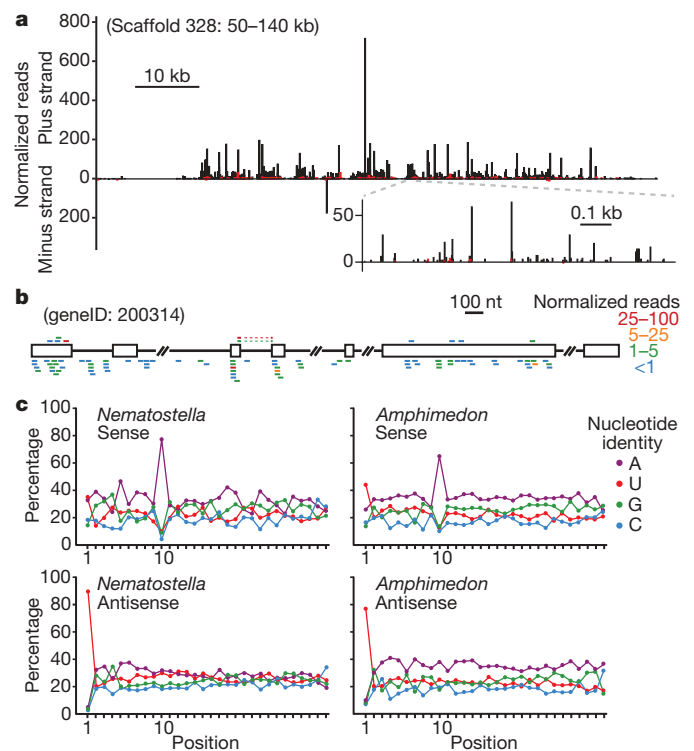


Figure 4 | The piRNAs of basal metazoans. **a**, Distribution of reads matching a *Nematostella* piRNA locus. Plotted is the number of matching reads with 5' nucleotide falling within each 100-nucleotide window (main graph) or at each nucleotide (higher-resolution inset) spanning the genomic region. Bars above and below the x axis indicate matches to the indicated strand, with black bars indicating reads with a 5'-U and red bars indicating the sum of all other reads. For reads also matching other genomic loci, counts were normalized by total genome matches. Other annotated piRNA loci are presented in Supplementary Tables 3 and 4. kb, kilobases. **b**, An annotated pre-mRNA corresponding to numerous small RNAs resistant to periodate treatment. Annotated coding segments (open boxes) and intron segments (black line) are indicated. The gene was homologous to endonuclease/reverse transcriptases of other genomes and presumed to be a transposase. Small RNAs with unique 5' ends are represented by coloured bars above or below the transcript (sense and antisense, respectively), with colours indicating the read numbers (normalized to account for the number of transcriptome matches). Small RNAs matching splice junctions (observed only for sense reads) are represented by discontinuous bars, linked by dashed lines. Other *Nematostella* and *Amphimedon* coding regions matching candidate piRNAs are listed in Supplementary Tables 5 and 6. **c**, Nucleotide composition of periodate-resistant small RNAs matching the indicated strand of *Nematostella* or *Amphimedon* annotated coding regions.

position 10. Examination of all 27–30-nucleotide periodate-resistant reads antisense to *Nematostella* coding regions revealed a propensity for a 5'-U, characteristic of primary piRNAs (Fig. 4c). The sense-strand piRNAs lacked this 5'-U bias and instead displayed a propensity for an A at position 10 (Fig. 4c and Supplementary Fig. 5). Moreover, sense and antisense reads that paired to each other tended to have 10 base pairs formed between their 5' ends (Supplementary Fig. 6). For the 24–30-nucleotide periodate-resistant reads from *Amphimedon*, the same hallmark features of the back-and-forth, or ping-pong, amplification cycle for piRNA biogenesis^{4,25} were observed (Fig. 4c and Supplementary Fig. 6). We conclude that the two classes of piRNAs found previously in mammals and flies have existed since the origin of metazoans: the class I piRNAs, represented by the mammalian pachytene piRNAs, which have unknown function during germline development, and the class II piRNAs, which use the ping-pong cleavage and amplification cascade to quiet expression of certain genes, particularly those of transposons. Indeed, the sequence-based transposon silencing by piRNAs, which by virtue of the feed-forward amplification process focuses on the most active transposon species, might be one of the principle drivers of transposon diversity in animals.

Taken together, our results indicate that miRNAs and piRNAs, as classes of small riboregulators, have been present since the dawn of animal life, and indeed might have helped to usher in the era of multicellular animal life. However, metazoan miRNA evolution seems to have been very dynamic: all miRNAs have been lost in *Trichoplax*, and the pre-miRNAs of Porifera, Cnidaria and Bilateria have assumed distinct sizes. In addition, no miRNAs have recognizable conservation between poriferans, cnidarians and bilaterians, with only one of the *Nematostella* miRNAs displaying recognizable homology to bilaterian miRNAs, either because it is the only homologue of extant bilaterian miRNAs or because divergence has obscured common ancestry of other miRNAs. The wholesale shifts in miRNA function implied by this plasticity are congruent with the report that, although thousands of miRNA–target interactions have been maintained within each of the nematode, fly and vertebrate lineages, very few appear to be conserved throughout all three lineages²⁶. The plasticity of miRNA sequences over long timescales helps to explain why the rich small-RNA biology in basal organisms had escaped detection for so long.

METHODS SUMMARY

The *M. brevicollis* library was constructed as described¹⁴ and sequenced by 454 Life Sciences. All other libraries (Supplementary Table 7) were constructed using an analogous method and sequenced on the Illumina platform.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 5 June; accepted 12 September 2008.

Published online 1 October 2008.

1. Cerutti, H. & Casas-Mollano, J. A. On the origin and functions of RNA-mediated silencing: from protists to man. *Curr. Genet.* **50**, 81–99 (2006).
2. Bartel, D. P. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116**, 281–297 (2004).
3. Lewis, B. P., Burge, C. B. & Bartel, D. P. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* **120**, 15–20 (2005).
4. Brennecke, J. *et al.* Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell* **128**, 1089–1103 (2007).
5. Aravin, A. A., Hannon, G. J. & Brennecke, J. The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race. *Science* **318**, 761–764 (2007).
6. Jones-Rhoades, M. W., Bartel, D. P. & Bartel, B. MicroRNAs and their regulatory roles in plants. *Annu. Rev. Plant Biol.* **57**, 19–53 (2006).
7. Molnar, A., Schwach, F., Studholme, D. J., Thuenemann, E. C. & Baulcombe, D. C. miRNAs control gene expression in the single-cell alga *Chlamydomonas reinhardtii*. *Nature* **447**, 1126–1129 (2007).

8. Zhao, T. *et al.* A complex system of small RNAs in the unicellular green alga *Chlamydomonas reinhardtii*. *Genes Dev.* **21**, 1190–1203 (2007).
9. Pasquinelli, A. E. *et al.* Conservation of the sequence and temporal expression of *let-7* heterochronic regulatory RNA. *Nature* **408**, 86–89 (2000).
10. Hertel, J. *et al.* The expansion of the metazoan microRNA repertoire. *BMC Genomics* **7**, 25 (2006).
11. Sempere, L. F., Cole, C. N., McPeck, M. A. & Peterson, K. J. The phylogenetic distribution of metazoan microRNAs: insights into evolutionary complexity and constraint. *J. Exp. Zool.* **306**, 575–588 (2006).
12. Prochnik, S. E., Rokhsar, D. S. & Aboobaker, A. A. Evidence for a microRNA expansion in the bilaterian ancestor. *Dev. Genes Evol.* **217**, 73–77 (2007).
13. Putnam, N. H. *et al.* Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. *Science* **317**, 86–94 (2007).
14. Ruby, J. G. *et al.* Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in *C. elegans*. *Cell* **127**, 1193–1207 (2006).
15. Ruby, J. G. *et al.* Evolution, biogenesis, expression, and target predictions of a substantially expanded set of *Drosophila* microRNAs. *Genome Res.* **17**, 1850–1864 (2007).
16. Larroux, C. *et al.* Genesis and expansion of metazoan transcription factor gene classes. *Mol. Biol. Evol.* **25**, 980–996 (2008).
17. Srivastava, M. *et al.* The *Trichoplax* genome and the nature of placozoans. *Nature* **454**, 955–960 (2008).
18. Lee, Y., Han, J., Yeom, K. H., Jin, H. & Kim, V. N. Drosha in primary microRNA processing. *Cold Spring Harb. Symp. Quant. Biol.* **71**, 51–57 (2006).
19. Fukuda, T. *et al.* DEAD-box RNA helicase subunits of the Drosha complex are required for processing of rRNA and a subset of microRNAs. *Nature Cell Biol.* **9**, 604–611 (2007).
20. King, N. *et al.* The genome of the choanoflagellate *Monosiga brevicollis* and the origin of metazoans. *Nature* **451**, 783–788 (2008).
21. Yao, M.-C. & Chao, J.-L. RNA-guided DNA deletion in *Tetrahymena*: an RNAi-based mechanism for programmed genome rearrangements. *Annu. Rev. Genet.* **39**, 537–559 (2005).
22. Horwich, M. D. *et al.* The *Drosophila* RNA methyltransferase, DmHen1, modifies germline piRNAs and single-stranded siRNAs in RISC. *Curr. Biol.* **17**, 1265–1272 (2007).
23. Seitz, H., Ghildiyal, M. & Zamore, P. D. Argonaute loading improves the 5' precision of both microRNAs and their miRNA strands in flies. *Curr. Biol.* **18**, 147–151 (2008).
24. Aravin, A. A., Sachidanandam, R., Girard, A., Fejes-Toth, K. & Hannon, G. J. Developmentally regulated piRNA clusters implicate MILI in transposon control. *Science* **316**, 744–747 (2007).
25. Gunawardane, L. S. *et al.* A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in *Drosophila*. *Science* **315**, 1587–1590 (2007).
26. Chen, K. & Rajewsky, N. Deep conservation of microRNA–target relationships and 3' UTR motifs in vertebrates, flies, and nematodes. *Cold Spring Harb. Symp. Quant. Biol.* **71**, 149–156 (2006).
27. Yigit, E. *et al.* Analysis of the *C. elegans* Argonaute family reveals that distinct Argonautes act sequentially during RNAi. *Cell* **127**, 747–757 (2006).
28. Boulard, S. J., Nielsen, C., Economou, A. D. & Telford, M. J. Testing the new animal phylogeny: a phylum level molecular analysis of the animal kingdom. *Mol. Phylogenet. Evol.* **49**, 23–31 (2008).
29. Griffiths-Jones, S., Saini, H. K., van Dongen, S. & Enright, A. J. miRBase: tools for microRNA genomics. *Nucleic Acids Res.* **36**, D154–D158 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank M. Abedin and E. Begovic for preparing the *Monosiga* and *Trichoplax* samples, respectively, W. Johnston for technical assistance, and J. Grenier, C. Mayr, C. Jan and N. Lau for discussions. This work was supported by an NIH postdoctoral fellowship (A.G.), and by grants from the NIH (D.P.B.), Richard Melmon (M.S., N.K. and D.S.R.), the Center for Integrative Genomics (M.S. and D.S.R.), the Gordon and Betty Moore Foundation (N.K.) and the Australian Research Council (B.F., B.J.W. and B.M.D.). D.P.B. is an investigator of the Howard Hughes Medical Institute.

Author Contributions A.G. constructed the libraries using procedures developed by H.R.C., and analysed the sequencing reads and protein homology. M.S., B.F., B.J.W., N.K., B.M.D. and D.S.R. provided samples for RNA extraction. A.G. and D.P.B. designed the study and prepared the manuscript, with input from other authors.

Author Information RNA sequencing data were deposited in the Gene Expression Omnibus (<http://www.ncbi.nlm.nih.gov/geo/>) under accession number GSE12578. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to D.P.B. (dbartel@wi.mit.edu).

METHODS

Small RNA sequencing. Samples of *N. vectensis* (mixed developmental stages, including adult), *A. queenslandica* (adult tissue, stored in RNAlater, Ambion) and *M. brevicollis* were ground under liquid nitrogen, and then RNA was extracted with Trizol (Invitrogen). RNA from *T. adhaerens* (mixed developmental stages, including adult) and *A. queenslandica* (mixed embryos, from cleavage stage to the larval stage³⁰, stored in RNAlater) was extracted directly with Trizol. The *M. brevicollis* library was constructed as described¹⁴ and sequenced by 454 Life Sciences. All other libraries (Supplementary Table 7) were sequenced on the Illumina platform, and prepared as follows. The 18–30-nucleotide RNAs were purified from total RNA (typically 5 µg) using denaturing polyacrylamide–urea gels. Before purification, trace amounts of 5′-³²P-labelled RNA size markers (AGCGUGUAGGGAUCCAAA and GGCAUUAACGCGGCCGCUCUACAAUAGUGA) were mixed with the total RNA and used to monitor this purification and subsequent ligations and purifications. The gel-purified RNA was ligated to pre-adenylated adaptor DNA (AppTCGTATGCCGTC-TTCTGCTTG-[3′-3′ linkage]-T) using T4 RNA ligase (10 units ligase, GE Healthcare, 10 µl reaction, 50 pmol adaptor ATP-free ligase buffer³¹, for 2 h at 21–23 °C). Gel-purified ligation products were ligated to a 5′-adaptor RNA (GUUCAGAGUUCUACAGUCCGACGAUC), again using T4 RNA ligase (as above, except with 20 units ligase, 15 µl reaction supplemented with 4 nmol ATP, 400 pmol adaptor, for 18 h at room temperature). Gel-purified ligation products were reverse-transcribed (SuperScript II, Invitrogen, 30 µl reaction with the reverse transcription primer CAAGCAGAAGACGGCATCA) and then RNA was base-hydrolysed with addition of 5 µl of 1 M NaOH and incubation at 90 °C for 10 min, followed by neutralization with addition of 25 µl 1 M HEPES, pH 7.0, and desalting (Microspin G-25 column, Amersham). The resulting cDNA library was amplified with the RT primer and PCR primer (AATGATACGGCGACCCAGCAGGTTCTCAGTCTTACAGTCCGA) for a sufficient number of cycles (typically ~20) to detect (SYBR Gold, Invitrogen) a clear band in a 90% formamide, 8% acrylamide gel, used for purification. Gel-purified amplicon (85–105 nucleotides) from each library was subjected to Illumina sequencing. The adaptor and primer sequences enabled cluster generation on the Illumina machine and placed a binding site for the sequencing primer (CGACAGGTTCTCAGGTTCTCAGTCCGACGATC) adjacent to the sequence of the small RNA. Periodate-treated libraries were generated identically, except total RNA was first subjected to β-elimination³². Mock-treated libraries omitting periodate were constructed in parallel.

MicroRNA identification and analysis. The *N. vectensis*, *T. adhaerens* and *M. brevicollis* genomes and predicted gene sets^{13,17,20} were downloaded from JGI (<http://jgi.doe.gov>); the *A. queenslandica* genome was a preliminary assembly¹⁶. After removing the adaptor sequences, reads were collapsed to a non-redundant set and matched to the appropriate genome. Genome matches were clustered if neighbouring matches fell within either 50 nucleotides (*Amphimedon*, *Nematostella*) or 500 nucleotides (*Amphimedon*) of each other. The increased size of the clustering window used for the *Amphimedon* analysis (500 nucleotides) was necessary because the 50-nucleotide window was insufficient to identify all *Amphimedon* miRNAs, owing to the increased size of their pre-miRNAs (Fig. 3e). No additional miRNAs were identified in *Nematostella* when using a 500-nucleotide window. Sequences of clusters containing 17–25-nucleotide reads cloned at least twice were folded with RNAfold³³. If the most frequently sequenced species was located on one arm of a predicted hairpin and the region of the hairpin corresponding to that sequence contained ≥16 base pairs, the candidate locus was examined manually for characteristics of known miRNAs, using criteria described in the main text. Before comparing between adult and embryonic libraries (Fig. 3d), counts corresponding to each mature miRNA from each library were first normalized by the total number of genome-matching reads in that library.

To detect possible homology between previously known miRNAs and either *Nematostella* or *Amphimedon* miRNAs, we searched miRBase (version 10.1) for miRNAs similar to our new miRNAs. Because miRNA conservation is most pronounced within the miRNA 5′ region³⁴, we first identified any known and new miRNAs that shared a hexanucleotide within their first eight nucleotides, allowing two-nucleotide offsets. Because of the limited length of the search sequence, and the large number of miRNAs in miRBase, most *Nematostella* or

Amphimedon miRNAs shared a hexanucleotide with miRBase miRNAs. For all such cases, we then searched for extended similarity between the pairs of miRNAs. With the exception of the miR-100 relationship, no more than chance similarity was observed (Supplementary Fig. 1). However, we cannot rule out the possibility that additional homologous relationships are present but undetectable. Because miRNAs are shorter than most other genetically encoded molecules, sequence divergence can more easily obscure homologous relationships, and although they resist changes in the seed region, which is crucial for target recognition, divergence in this 5′ region can be accelerated with the processes of sub- and neo-functionalization¹⁵.

Piwi-interacting RNA identification and analysis. *Nematostella* 27–30-nucleotide RNAs and *Amphimedon* 24–30-nucleotide RNAs were mapped to their respective genome, and at each matching locus counts were normalized, dividing by the number of genome matches for the sequenced RNA. Regions with both a high number of match-normalized reads (*Nematostella*: >1,000 per 10 kilobases; *Amphimedon*: >100 per 5 kilobases) and a high diversity of read sequences (*Nematostella*: >500 different sequences per 10 kilobases; *Amphimedon*: >50 different sequences per 5 kilobases) were identified; following the periodate experiment we further evaluated these regions, which led to the removal of four *Amphimedon* regions that had far fewer reads in the periodate-treated libraries. The remaining regions are listed in Supplementary Tables 3 (*Nematostella*) and 4 (*Amphimedon*), which report the proportion of 5′-U match-normalized reads to each strand and the ratio of match-normalized read counts in periodate-treated compared to mock-treated libraries, after normalization for the number of genome-matching reads in each library. The number of predicted transcripts^{13,16} overlapping genomic piRNA clusters (Supplementary Tables 3 and 4) was calculated and compared to the number overlapping 1,000 random sets equal in size and number to the piRNA clusters. Inferred protein sequences from predicted transcripts matching the greatest number of periodate-resistant, match-normalized reads were compared to annotated protein sequences using BLAST. Transcripts that were significantly similar to annotated transposons, or protein domains implicated as transposases (for example reverse transcriptases) were considered to encode transposases. A random selection of 100 predicted transcripts was searched similarly to ascertain significance (*Nematostella*: 3 out of 100; *Amphimedon*: 6 out of 100). When mapping to annotated protein-coding regions (Fig. 4b), reads with both sense and antisense matches were distributed to both the sense and antisense tallies after weighting by the proportion of their sense and antisense matches.

Cataloguing of the small RNA machinery. To identify homologues of components of the small RNA machinery, all established family members from *H. sapiens*, *D. melanogaster*, *C. elegans*, *S. pombe* and *A. thaliana* were used as BLAST query sequences against all annotated protein sequences of each species in Table 1. The top-ranking hits resulting from these initial searches were used reciprocally as query sequences against all annotated protein sequences of *H. sapiens*, *D. melanogaster*, *C. elegans*, *S. pombe* and *A. thaliana*. If the top-ranking hits of such reciprocal queries corresponded to an established family member, the query sequence was considered to be a candidate homologue. The domain structure of each candidate sequence was then evaluated³⁵, and candidates lacking the diagnostic domains were discarded. The diagnostic domains used were a Paz and a Piwi domain (for Ago and Piwi family members), two RNase III domains (Dicer and Drosha), a double-stranded RNA-binding domain (Pasha) and a methylase domain (Hen1).

- Adamska, M. et al. Wnt and TGF-β expression in the sponge *Amphimedon queenslandica* and the origin of metazoan embryonic patterning. *PLoS ONE* 2, e1031 (2007).
- England, T. E., Gumport, R. I. & Uhlenbeck, O. C. Dinucleoside pyrophosphate are substrates for T4-induced RNA ligase. *Proc. Natl Acad. Sci. USA* 74, 4839–4842 (1977).
- Kemper, B. Inactivation of parathyroid hormone mRNA by treatment with periodate and aniline. *Nature* 262, 321–323 (1976).
- Hofacker, I. L. Fast folding and comparison of RNA secondary structures. *Monatsh. Chem.* 125, 167–188 (1994).
- Lim, L. P. et al. The microRNAs of *Caenorhabditis elegans*. *Genes Dev.* 17, 991–1008 (2003).
- Marchler-Bauer, A. et al. CDD: a conserved domain database for interactive domain family analysis. *Nucleic Acids Res.* 35, D237–D240 (2007).

ARTICLES

Activity-dependent regulation of inhibitory synapse development by Npas4

Yingxi Lin¹, Brenda L. Bloodgood¹, Jessica L. Hauser^{1†}, Ariya D. Lapan², Alex C. Koon^{1‡}, Tae-Kyung Kim¹, Linda S. Hu¹, Athar N. Malik^{1,3} & Michael E. Greenberg¹

Neuronal activity regulates the development and maturation of excitatory and inhibitory synapses in the mammalian brain. Several recent studies have identified signalling networks within neurons that control excitatory synapse development. However, less is known about the molecular mechanisms that regulate the activity-dependent development of GABA (γ -aminobutyric acid)-releasing inhibitory synapses. Here we report the identification of a transcription factor, Npas4, that plays a role in the development of inhibitory synapses by regulating the expression of activity-dependent genes, which in turn control the number of GABA-releasing synapses that form on excitatory neurons. These findings demonstrate that the activity-dependent gene program regulates inhibitory synapse development, and suggest a new role for this program in controlling the homeostatic balance between synaptic excitation and inhibition.

Sensory experience controls multiple steps in the development and maturation of synapses in the mammalian brain^{1–4}. Many of the effects of neuronal activity are mediated by the release of glutamate at excitatory synapses and the subsequent influx of calcium (Ca^{2+}) into the postsynaptic neuron. This results in changes in the number and strength of synapses, a process that underlies learning and memory as well as animal behaviour.

Neurons in the central nervous system receive excitatory synaptic input from glutamatergic neurons and inhibitory input from GABA-releasing (GABAergic) interneurons, except during early development when the first GABAergic synapses are depolarizing and provide the excitatory drive critical for the subsequent development of glutamatergic synapses⁵. The proper balance between excitatory and inhibitory synapses is crucial for representation of sensory information^{6,7}, execution of motor commands^{8,9} and higher-order cognitive functions¹⁰. Neurological disorders such as autism, schizophrenia and epilepsy are associated with an imbalance between excitatory and inhibitory synapses^{11–13}. The number or strength of excitatory synapses can be modified in response to changes in activity, and the molecular mechanisms of these processes have been extensively investigated^{14–16}. Less is known about the activity-dependent regulation of inhibitory synapses.

The density of inhibitory synapses in brain regions such as primary sensory cortex, hippocampus and cerebellum is regulated by the level of excitatory synaptic activity and sensory input^{17–22}. In addition, initiation of the critical period for synaptic plasticity in the visual cortex is dependent on visual activity and strongly influenced by the maturation of inhibitory synapses²³, suggesting that the activity-dependent regulation of GABAergic synapses is important for the plasticity of the nervous system. Finally, recent studies indicate that regulation of GABAergic synapses in response to neuronal activity may be a critical component of the homeostatic mechanism that maintains a balance between excitation and inhibition in the face of fluctuations in the level of sensory input into neural circuits²⁴. Despite the accumulating evidence that neuronal activity regulates

the development and maintenance of inhibitory synapses, the molecular mechanisms that control these processes remain to be characterized.

Here we identify a transcription factor, Npas4, that is critical for activity-dependent regulation of GABAergic synapse development. Npas4 expression is rapidly activated by excitatory synaptic activity and turns on a program of gene expression that triggers the formation and/or maintenance of inhibitory synapses on excitatory neurons. These findings provide a molecular link between neuronal excitation and GABAergic synapse development, and suggest a new role for the activity-dependent gene program in controlling inhibitory synapse formation/maintenance on excitatory neurons.

Npas4 is regulated by neuronal activity

The formation of inhibitory synapses onto excitatory neurons is regulated by neuronal activity, takes place over several days, and is a cell-wide process that results in the formation of synapses onto both the cell body and dendrites^{18,25}. These features led us to hypothesize that activity-dependent development of inhibitory synapses might be controlled postsynaptically by one or more activity-regulated genes. To test this hypothesis, we used DNA microarrays to identify genes that are induced by membrane depolarization in mouse cortical neurons at the time when inhibitory synapses are developing.

We identified more than 300 genes whose expression levels were altered upon membrane depolarization (Gene Expression Omnibus accession number GSE11256), a third of which were novel activity-regulated genes not seen in previous screens^{26,27}. We looked for genes predicted to encode transcription factors, reasoning that, through genome-wide characterization of the targets of an activity-regulated transcription factor that controls inhibitory synapse number, we could gain insight into the biological program that is important for inhibitory synapse development. Among the approximately 20 known or putative transcription factors identified, we focused on genes that are selectively induced by Ca^{2+} influx in neurons but not other cell types, that are transcribed in response to excitatory

¹F. M. Kirby Neurobiology Center, Children's Hospital and Departments of Neurology and Neurobiology, Harvard Medical School, 300 Longwood Avenue, Boston, Massachusetts 02115, USA. ²Program in Biological and Biomedical Sciences, Harvard Medical School, 240 Longwood Avenue, Boston, Massachusetts 02115, USA. ³Program in Neuroscience, Harvard Medical School, 300 Longwood Avenue, Boston, Massachusetts 02115, USA. [†]Present addresses: Baylor College of Medicine, Medical Scientist Training Program, One Baylor Plaza Suite N201, MS:BCM215, Houston, Texas 77030-7498, USA (J.L.H.); University of Massachusetts Medical School, Lazare Medical Research Building, Room 760C, 364 Plantation Street, Worcester, Massachusetts 06105, USA (A.C.K.).

synaptic activity, and that are expressed coincidentally with the development of inhibitory synapses. One transcription factor, the bHLH-PAS family member Npas4 (refs 28–31), fulfilled all these criteria (Fig. 1) and was investigated further.

Unlike other activity-dependent transcription factors such as CREB and c-fos, Npas4 expression in neurons is selectively induced by Ca^{2+} influx but not by several neurotrophic factors, growth factors or forskolin, an activator of protein kinase A (Fig. 1b). Furthermore, Npas4 induction is transient (Fig. 1c), occurs selectively in neurons, and predominately in excitatory neurons (data not shown). Npas4 expression is barely detectable in immature primary hippocampal neurons, when there are few synapses³², and increases with the formation and maturation of synapses²⁵ (Fig. 1d).

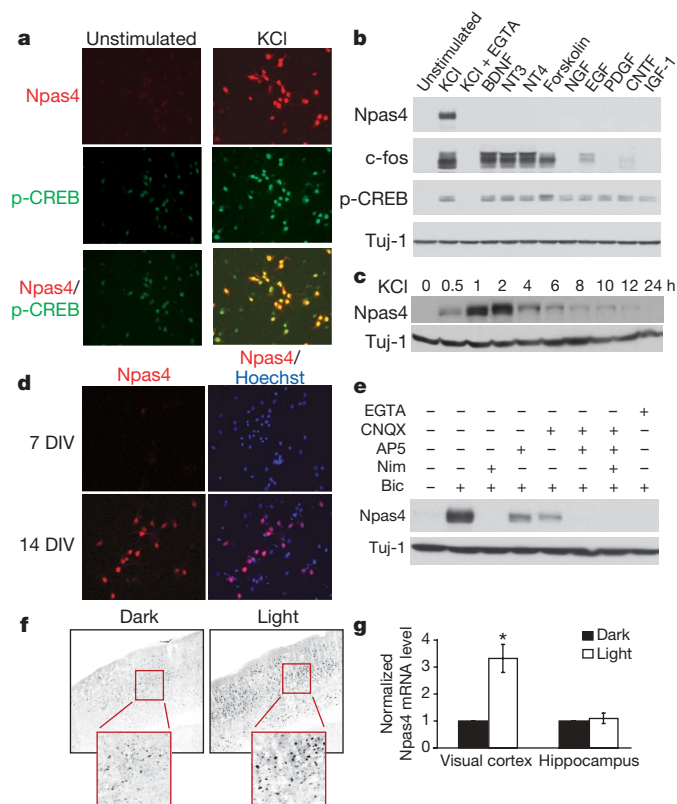


Figure 1 | Npas4 expression is regulated by neuronal activity *in vitro* and *in vivo*. **a**, Immunostaining showing Npas4 protein is induced in rat hippocampal neurons (7 DIV) by depolarization (50 mM KCl, 2 h, right). p-CREB, CREB phosphorylated at Ser 133. **b**, Western blot showing Npas4 is selectively induced by membrane depolarization (50 mM KCl, 7 DIV rat hippocampal neurons), but not by BDNF (50 ng ml⁻¹), NT3 (50 ng ml⁻¹), NT4 (50 ng ml⁻¹), forskolin (10 μM), NGF (100 ng ml⁻¹), EGF (100 ng ml⁻¹), PDGF (100 ng ml⁻¹), CNTF (100 ng ml⁻¹) or IGF-1 (100 ng ml⁻¹). Induction is prevented by pretreatment with the Ca^{2+} chelator EGTA (5 mM, 10 min). **c**, Western blot showing Npas4 (7 DIV rat hippocampal neurons) is transiently induced by membrane depolarization (50 mM KCl, 30 min). **d**, Basal Npas4 expression increases as neurons mature, presumably because of increased endogenous spontaneous activity: compare immunostaining of 7 and 14 DIV rat hippocampal neurons. **e**, Western blot showing stimulation of primary hippocampal neurons (14 DIV) with bicuculline (50 μM, 2 h) increases Npas4 expression levels. This is prevented by pretreatment with nimodipine (5 μM, 1 h) or EGTA (5 μM, 5 min) and reduced by pretreatment (1 h) with antagonists to NMDA receptors (100 μM AP5) or AMPA receptors (50 μM CNQX). **f**, **g**, Mice dark-reared for one week (P21–P28) and then stimulated with strobe lights have greater Npas4 expression levels in the visual cortex than their dark-reared littermates, but there is no difference in the hippocampus. Light stimulation was applied for 2 h for immunocytochemistry analysis (**f**) or 1 h for Npas4 mRNA quantification (**g**). Significance was determined using a one-tailed paired *t*-test, **P* < 0.05. Data are shown as mean ± s.e.m.

Npas4 is induced in cultured neurons by the GABA_A-receptor antagonist bicuculline, which increases action-potential firing and excitatory synaptic transmission (Fig. 1e). This induction requires an influx of extracellular Ca^{2+} through L-type voltage-sensitive calcium channels (L-VSCCs) and is partly dependent on the activation of *N*-methyl-D-aspartate (NMDA) and α -amino-3-hydroxy-5-methyl-4-isoxazole propionic acid (AMPA) receptors (Fig. 1e). Npas4 expression is also induced in pertinent brain regions *in vivo* in response to specific stimuli: visual stimulation of mice after a period of dark-rearing results in an increase in Npas4 messenger RNA (mRNA) and protein levels specifically in the visual cortex (Fig. 1f, g).

Npas4 regulates the development of inhibitory synapses

The effect of Npas4 on inhibitory synapse development was investigated by RNA interference (RNAi)-mediated knockdown in excitatory neurons of rat dissociated hippocampal cultures, before synapse formation was underway. A small hairpin RNA targeting Npas4 (Npas4-RNAi) effectively reduced Npas4 expression (Fig. 2a) without affecting the overall health of the neurons (Supplementary Fig. 1). To measure inhibitory synapse number, neurons were immunostained for presynaptic GABA-producing enzyme GAD65 and postsynaptic GABA_A-receptor $\gamma 2$ subunit (GABA_A- $\gamma 2$). Co-localization of GAD65 and GABA_A- $\gamma 2$ puncta on a green fluorescent protein (GFP)-transfected glutamatergic neuron was considered indicative of a synapse (Fig. 2b). Expression of Npas4-RNAi, but not a scrambled RNAi (control-RNAi), significantly reduced the number of inhibitory synapses (Fig. 2c). Analysis of synapse number using antibodies that recognize a second pair of inhibitory synaptic proteins (GAD67 and GABA_A- $\beta 2/3$) gave a similar result (data not shown). These data suggest that Npas4 positively regulates the number of inhibitory synapses that form on excitatory neurons.

Different classes of inhibitory neurons synapse onto distinct perisomatic or dendritic regions of pyramidal neurons³³. We found that Npas4-RNAi leads to a reduction in the number of inhibitory synapses formed on both the perisomatic and dendritic regions, suggesting that Npas4 regulates the number of inhibitory synapses formed by multiple classes of inhibitory neurons (Fig. 2c). In both regions, Npas4-RNAi significantly reduced the density of postsynaptic GABA_A- $\gamma 2$ puncta but had less effect on presynaptic GAD65 puncta, compared with the control-RNAi (Fig. 2d). These findings suggest that Npas4 regulates inhibitory synapse number by controlling the number of postsynaptic specializations, resulting in subsequent remodelling or retraction of the presynaptic terminals.

To test whether Npas4 is important for the development of functional inhibitory synapses in a more intact neural circuit, rat organotypic hippocampal cultures were biolistically co-transfected with GFP and either a control vector or Npas4-RNAi, and whole-cell recordings were performed on GFP-positive CA1 pyramidal neurons to measure spontaneous miniature inhibitory postsynaptic currents (mIPSCs, Fig. 3a). Knocking down Npas4 expression in organotypic cultures significantly increased the inter-event interval and decreased the amplitude of mIPSCs (Fig. 3b, c). This effect of Npas4-RNAi in organotypic slices may be indicative of a decrease in inhibitory synapse number, consistent with our observation that Npas4-RNAi reduces inhibitory synapse number in dissociated cultures. Together, these findings indicate that Npas4 plays an important role in regulating the number of functional inhibitory synapses received by an excitatory neuron.

Consistent with a role for Npas4 in inhibitory synapse development, Npas4 knockout mice (Npas4^{-/-}, Supplementary Fig. 2) appear anxious and hyperactive, are prone to seizures and have a shortened lifespan compared with their wild-type littermates, phenotypically resembling other knockout mice lacking genes that control the formation or function of inhibitory circuits^{34–36}. However, we found that the frequencies of mIPSCs in acute hippocampal slices prepared from wild-type and Npas4^{-/-} mice were similar (Supplementary Fig. 3), in contrast to the clear change in mIPSC frequency caused by disrupting Npas4 expression acutely in organotypic slices. This difference may

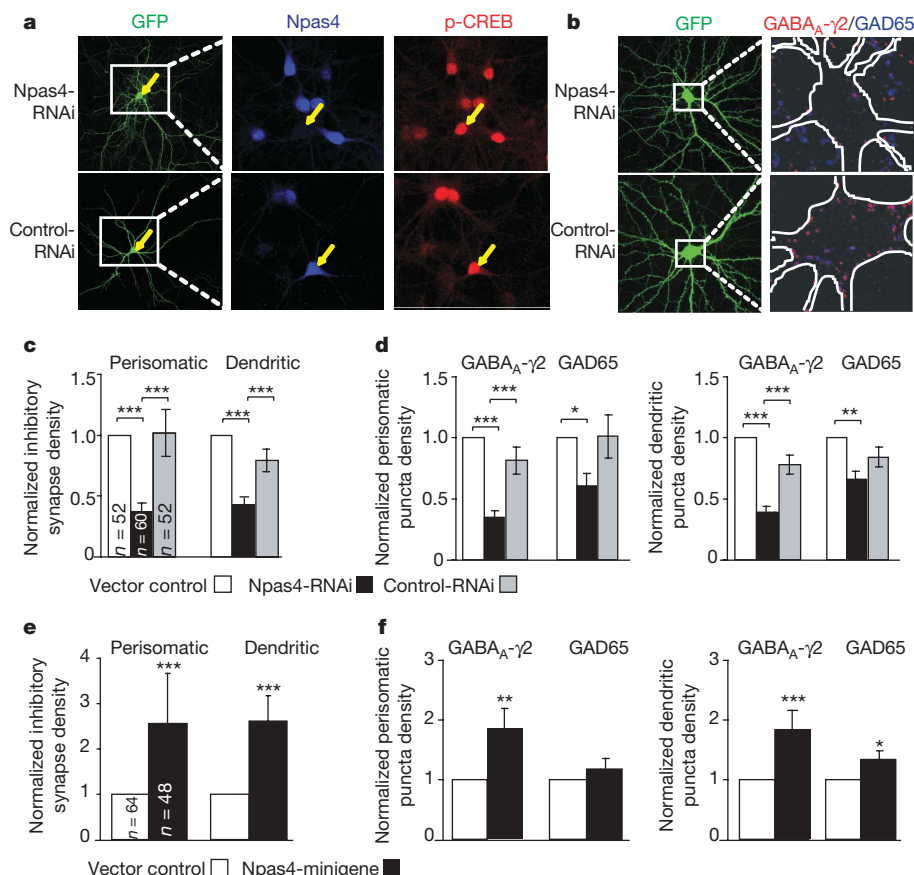


Figure 2 | Npas4 regulates the number of GABAergic synapses in cultured hippocampal neurons. **a**, Npas4-RNAi, but not control-RNAi, reduces the expression of Npas4 in primary hippocampal neurons. Cultures were transfected at 6 DIV and stimulated with bicuculline (50 μ M, 2 h) at 14 DIV. **b**, The number of GABAergic synapses is significantly reduced by Npas4-RNAi, as illustrated by two representative rat hippocampal neurons. Cultures were co-transfected (6 DIV) with GFP and either Npas4-RNAi (top) or control-RNAi (bottom). Cultures were subsequently immunostained (25 DIV) with antibodies against GAD65 (blue) and GABA_A-γ2 (red). **c**, Quantification of the normalized density of co-localized GABA_A-γ2 and

GAD65 puncta in 14 DIV rat hippocampal neurons transfected with vector control, Npas4-RNAi or control-RNAi. **d**, Separate quantification of perisomatic and dendritic GABA_A-γ2 and GAD65 puncta measured in **c**. **e**, Npas4-minigene increases the density of co-localized GABA_A-γ2 and GAD65 puncta. **f**, Separate quantification of perisomatic and dendritic GABA_A-γ2 and GAD65 puncta measured in **e**. See Methods for details of data normalization and error propagation. Significance was determined using multifactorial analysis of variance. * $P < 0.05$; ** $P < 0.005$; *** $P < 0.0005$. Data are presented as mean \pm s.e.m. from three (**c**, **d**) or four (**e**, **f**) independent experiments; total numbers of neurons analysed (n) are indicated.

be due to the activation of compensatory pathways during development in the absence of Npas4. Alternatively, it may reflect the fact that Npas4 is absent in both pre- and postsynaptic cells in acute slices from Npas4^{-/-} mice, whereas in the experiments with organotypic slices the disruption of Npas4 expression occurs in postsynaptic neurons only.

To determine conclusively the involvement of Npas4 in inhibitory synapse development, we generated an Npas4 conditional knockout mouse (Npas4^{flx/flx}) in which the coding region of Npas4 is flanked by loxP sites and can be acutely removed by Cre-mediated recombination (Supplementary Fig. 4). Organotypic hippocampal slices were prepared from Npas4^{flx/flx} mice and whole-cell recordings made from CA1 pyramidal neurons transfected with GFP and either a control vector or a vector encoding Cre recombinase. Cre expression had no effect on mIPSCs in wild-type neurons (Supplementary Fig. 5). However, compared with transfection with the control construct, transfection of Npas4^{flx/flx} neurons with Cre led to a significant increase in the mIPSC inter-event interval (Fig. 3d, e). This finding provides further evidence that CA1 pyramidal neurons lacking Npas4 receive fewer inhibitory synaptic inputs.

We next investigated whether the number of inhibitory synapses forming onto a cell is controlled by the amount of Npas4 expressed in response to excitatory stimuli. If this is the case, increasing the level of Npas4 should lead to an increase in the number of inhibitory synapses formed on the Npas4-expressing neuron. Additional copies of the Npas4 gene were introduced into neurons using an

Npas4-minigene cassette consisting of all Npas4 introns and exons as well as 5 kilobases (kb) of genomic sequence 5'- and 3'- to the coding region. We verified that the Npas4-minigene drives ectopic expression of Npas4 that is activity regulated and functions similarly to the endogenous gene (Supplementary Fig. 6). We found that significantly more inhibitory synapses were formed onto cultured hippocampal neurons expressing the Npas4-minigene than onto neurons transfected with a control vector (Fig. 2e), largely because of an increase in the number of postsynaptic GABA_A-γ2 puncta (Fig. 2f). Furthermore, expression of the Npas4-minigene in CA1 pyramidal neurons of hippocampal slices significantly decreases the mIPSC inter-event interval and significantly increases the amplitude (Fig. 3b, c), consistent with the presence of additional and stronger inhibitory synapses on neurons expressing higher levels of Npas4.

Effect of Npas4 on excitatory synapses

The appropriate balance between excitation and inhibition is critical for the function of neural circuits. To maintain this balance, changes in the number or strength of inhibitory synapses are often coupled to changes in excitatory synapses²⁴. We next asked whether excitatory synapse number or function is also affected when Npas4 expression is perturbed.

We first determined whether Npas4 regulates the number of excitatory synapses in dissociated hippocampal cultures. As before, neurons were transfected with either Npas4-RNAi or control constructs,

before synaptogenesis was underway and thus before the balance between excitation and inhibition was established. Cultures were immunostained for the presynaptic marker synapsin1 and the excitatory postsynaptic marker PSD95, and the numbers of co-localized synapsin1 and PSD95 puncta on transfected glutamatergic neurons were quantified. Neither the total numbers of excitatory synapses nor the individual numbers of pre- or postsynaptic markers changed significantly upon expression of Npas4-RNAi (Fig. 4a, b), under conditions that significantly decreased the number of inhibitory synapses. Likewise, expression of the Npas4-minigene had no effect on excitatory synapse number (Fig. 4c, d). Because perturbation of the level of Npas4 expression occurred before synaptic connections were established, these experiments suggest that Npas4 is not a major contributor to excitatory synaptogenesis.

We next examined whether Npas4 affects excitatory synapse function in organotypic hippocampal slices, where homeostatic mechanisms are known to control excitatory/inhibitory balance within neural circuits. Deletion of the Npas4 gene from neurons in Npas4^{flx/flx} organotypic hippocampal slices by Cre recombination significantly decreased the inter-event interval of spontaneous miniature excitatory postsynaptic currents (mEPSCs, Fig. 4e, f), whereas elevating the level of Npas4 with the Npas4-minigene significantly increased the inter-event interval and decreased the amplitude of mEPSCs (Fig. 4g, h), compared with control-transfected neurons. Thus, abolishing Npas4 expression in a pyramidal neuron increases, whereas enhancing the level of Npas4 decreases, the number and/or presynaptic release probability of excitatory synapses that form on

the neuron. Therefore, the net result of Npas4 activation in an intact neural circuit is an increase in synaptic inhibition and a decrease in excitation of a neuron. We conclude that Npas4 induction in response to increased excitatory input acts to reduce the level of activity, and therefore may function as a negative feedback mechanism to maintain the homeostatic balance between excitation and inhibition.

Npas4 regulates genes that control inhibitory synapse development

To uncover the program of gene expression controlled by Npas4, we acutely knocked down Npas4 expression in a high percentage of wild-type neurons using a lentivirus expressing Npas4-RNAi (Supplementary Fig. 7) and performed a second DNA microarray

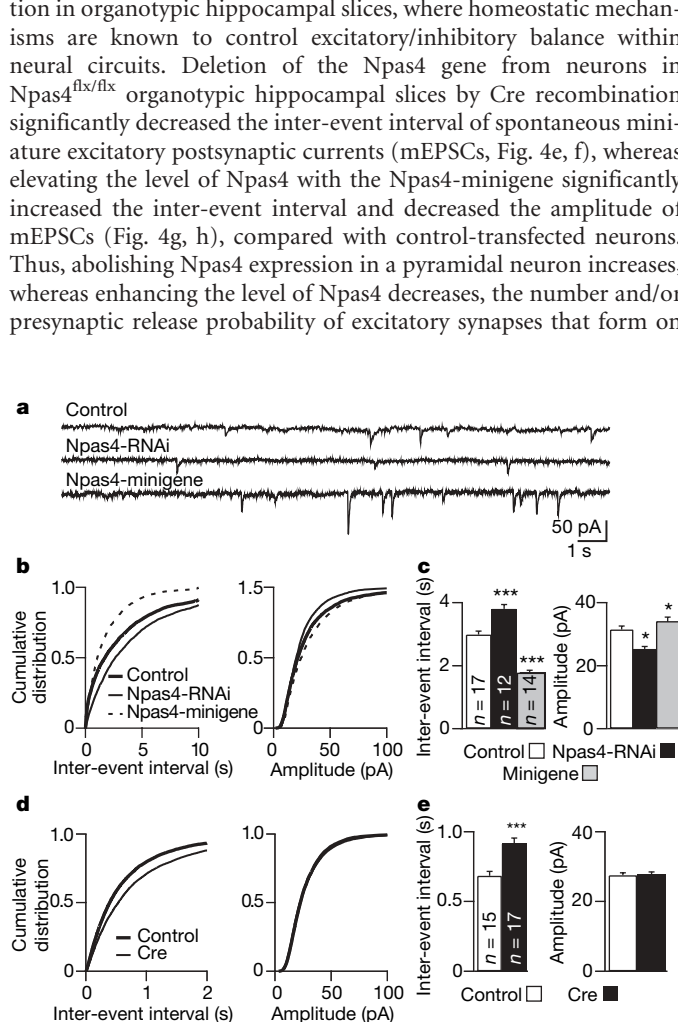


Figure 3 | Npas4 regulates GABAergic synapse development in organotypic hippocampal slices. **a**, Representative mEPSCs recorded from CA1 pyramidal neurons in organotypic hippocampal slices biolistically co-transfected with GFP and either vector control, Npas4-RNAi or Npas4-minigene. **b**, Cumulative distributions of mEPSC inter-event intervals and amplitudes recorded from neurons transfected with vector control, Npas4-RNAi or Npas4-minigene. **c**, Mean \pm s.e.m. of data from **b**. mEPSC inter-event intervals: 2986.3 \pm 105.7, 3803.0 \pm 136.9 and 1776.9 \pm 75.1 ms; amplitudes: 31.5 \pm 1.1, 25.7 \pm 0.8 and 34.1 \pm 1.3 pA; for vector control, Npas4-RNAi and Npas4-minigene, respectively. **d**, Cumulative distributions of mEPSC inter-event intervals and amplitudes recorded from Npas4^{flx/flx} neurons co-transfected with GFP and either vector control or Cre recombinase. **e**, Mean \pm s.e.m. of data from **d**. mEPSC inter-event intervals: 684.2 \pm 31.2 and 917.4 \pm 37.7 ms; amplitudes: 27.5 \pm 0.7 and 27.9 \pm 0.6 pA; for vector control and Cre, respectively. Total numbers of neurons analysed in each condition (*n*) are indicated in **c** and **e**. **P* < 0.05; ****P* < 0.001.

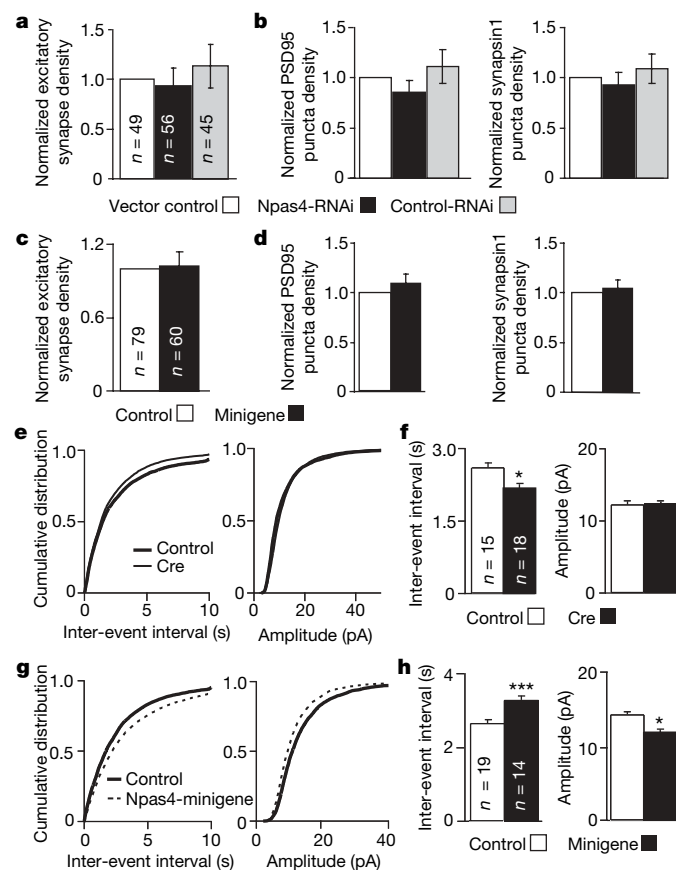


Figure 4 | Npas4 has no effect on excitatory synaptogenesis but affects excitatory/inhibitory balance in neural circuits. **a**, The number of excitatory synapses is not affected by Npas4-RNAi. Quantification of the normalized density of co-localized PSD95 and synapsin1 puncta in 14 DIV rat hippocampal neurons transfected with vector control, Npas4-RNAi or control-RNAi. **b**, Separate quantification of PSD95 and synapsin1 puncta measured in **a**. **c**, Npas4-minigene has no effect on the density of excitatory synapses. Quantification of co-localized synapsin1 and PSD95 puncta is shown. **d**, Separate quantification of PSD95 and synapsin1 puncta measured in **c**. In **a**–**d**, data are presented as mean \pm s.e.m. **e**, Cumulative distribution of mEPSC inter-event intervals and amplitudes recorded from Npas4^{flx/flx} neurons co-transfected with GFP and either vector control or Cre. **f**, Mean \pm s.e.m. of data from **e**. mEPSC inter-event intervals: 2581.4 \pm 104.1 ms and 2140.5 \pm 79.7 ms; amplitudes: 12.1 \pm 0.4 and 12.4 \pm 0.3 pA; for vector control and Cre, respectively. **g**, Cumulative distribution of mEPSC inter-event intervals and amplitudes recorded from neurons transfected with GFP and either vector control or Npas4-minigene. **h**, Mean \pm s.e.m. of data from **g**. mEPSC inter-event intervals: 2686.4 \pm 89.3 and 3320.7 \pm 117.2 ms; amplitudes: 14.3 \pm 0.3 and 12.0 \pm 0.3 pA; for vector control and Npas4-minigene, respectively. Total numbers of neurons analysed in each condition (*n*) are indicated. **a**, **b**, Three independent experiments; **c**, **d**, four independent experiments. **P* < 0.05; ****P* < 0.001.

experiment to identify activity-regulated genes that are misregulated in the absence of Npas4. The expression levels of 327 microarray probe sets representing 270 unique genes were significantly different in cultures expressing Npas4-RNAi compared with those expressing the control virus (Gene Expression Omnibus accession number GSE11258, Supplementary List). The expression of 182 of these genes was also acutely regulated by membrane depolarization in the absence of Npas4-RNAi, indicating that many putative Npas4-regulated genes are activity regulated (Fig. 5a). Although Npas4 has

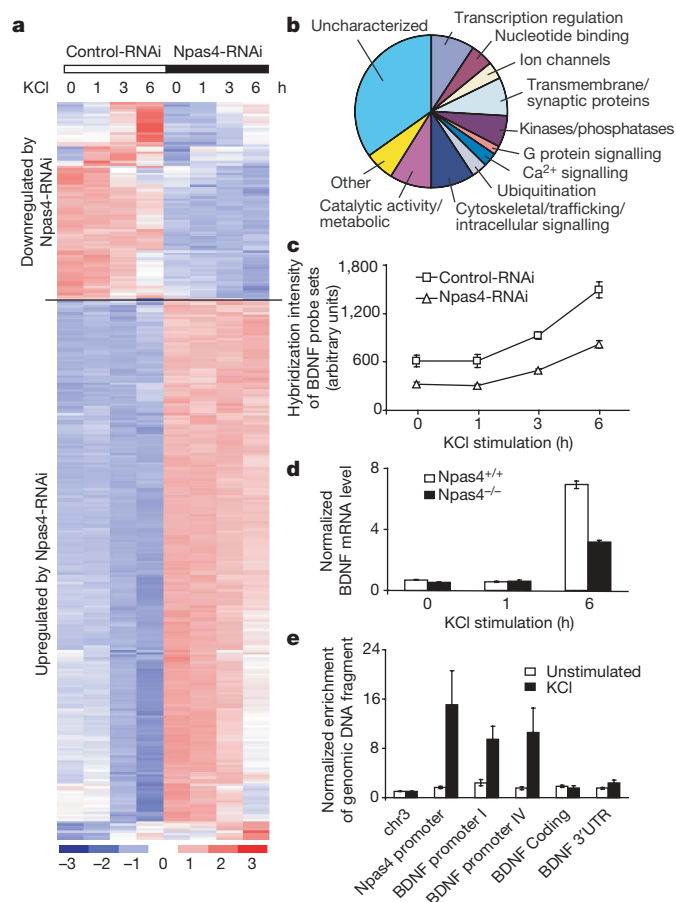


Figure 5 | Npas4 controls a program of gene expression that regulates GABAergic synapses. **a**, Hierarchical clustering of 327 probe sets (270 putative Npas4 target genes) based on their expression profiles using dChip⁴⁹. The expression level of each probe set is normalized to a mean of 0 and a standard deviation of 1. Expression values are displayed within the range [-3, 3] with levels above, equal to or below the mean displayed in red, white and blue, respectively. Dark red represents 3 or higher, and dark blue -3 or lower. **b**, Biological functions of 270 putative Npas4 target genes based on Gene Ontology information provided by Affymetrix (<http://www.affymetrix.com>). **c**, BDNF expression is reduced by Npas4-RNAi (1.9 ± 0.11 -fold reduction, $P < 0.01$, one-tailed paired t -test). Mean \pm s.e.m. from three independent experiments is shown, and each data point was generated by averaging the hybridization intensity of two BDNF probe sets (1422168_a_at and 1422169_a_at). **d**, BDNF levels are consistently reduced in neurons from Npas4^{-/-} mice compared with their wild-type littermates ($58.46 \pm 6.49\%$ decrease, 95% confidence interval 30.5–86.4%). Cortical cultures prepared from Npas4^{+/+} and Npas4^{-/-} littermates (7 DIV) were stimulated with KCl (55 mM), and BDNF mRNA levels were measured by quantitative reverse transcriptase PCR using primers in BDNF coding region. Three littermate pairs from three different litters were analysed; data (mean \pm s.e.m.) from one representative pair is shown. **e**, Npas4 interacts directly with BDNF promoters I and IV in an activity-dependent manner, as shown by chromatin immunoprecipitation. The Npas4 promoter is used as a positive control. Data are normalized to a control region on chromosome 3 and are presented as mean \pm s.e.m. from five independent experiments.

been shown to function as a transcriptional activator (Supplementary Fig. 6b)²⁸, we found that many genes are negatively regulated by Npas4 (Fig. 5a). This may reflect the fact that Npas4 functions as a transcriptional repressor as well as an activator, and/or that Npas4 indirectly affects gene expression by altering neuronal excitation.

Npas4 appears to regulate a wide variety of genes, such as activity-regulated immediate early genes, various classes of transcription factors, channel proteins, G-protein signalling molecules, kinases and phosphatases, and genes involved in pathways that modulate synaptic functions, such as ubiquitination, trafficking and receptor endocytosis (Fig. 5b and Supplementary List). Interestingly, the functions of 94 of the 270 putative Npas4-regulated genes are uncharacterized (Fig. 5b), suggesting that Npas4 regulates many genes that could affect inhibitory synapses in novel ways.

As a first step towards understanding the genetic program regulated by Npas4, we focused on targets that might be directly involved in the development of GABAergic synapses. Brain-derived neurotrophic factor (BDNF) stood out because it had previously been shown to regulate GABAergic synapse maturation and function^{18,20,37–39}. BDNF expression is consistently reduced by almost two-fold in cultures expressing Npas4-RNAi compared with control cultures (Fig. 5c). Primary cultures from Npas4^{-/-} mice showed a similar decrease in depolarization-induced BDNF expression compared with their wild-type littermates (Fig. 5d).

The BDNF gene has many promoters and the activity-dependent BDNF mRNA transcripts are controlled by promoters I and IV^{40–42}. Using a chromatin immunoprecipitation assay, we asked whether Npas4 binds to these promoters in primary cortical cultures that were left untreated or membrane-depolarized to trigger Npas4 expression. Although no Npas4 binding was detected in untreated neurons, in membrane-depolarized neurons Npas4 was found to be associated with BDNF promoters I and IV, but not with the coding region or 3' untranslated region (Fig. 5e), suggesting that Npas4 may directly regulate activity-dependent expression of BDNF.

To determine whether BDNF contributes to the effects of Npas4 on GABAergic synapse development, we tested whether the ability of the Npas4-minigene to increase the number of inhibitory synapses is attenuated by knockdown of BDNF expression, using a previously validated small hairpin RNA that targets the BDNF coding region (BDNF-RNAi)⁴³. Confirming that BDNF increases inhibitory synapse number⁴⁴, BDNF-RNAi increased mIPSC inter-event intervals compared with control values (2986.3 ± 109.7 and 4086.1 ± 140.7 ms, control and BDNF-RNAi, respectively, $P < 0.01$). Expression of the Npas4-minigene alone led to an approximately 40% decrease in the inter-event interval of mIPSCs recorded from CA1 neurons (Fig. 3b, c), as previously observed, but this was partly attenuated to approximately 20% by the presence of BDNF-RNAi (Fig. 6a). In addition, in the presence of BDNF-RNAi, the effect of the Npas4-minigene on the amplitude of mIPSCs was completely reversed (Fig. 6b). These findings suggest that BDNF mediates a portion of the effect of Npas4 on inhibitory synapse number, but that additional Npas4 targets may also be involved.

Discussion

We have identified the activity-regulated transcription factor Npas4 as a key regulator of GABAergic synapse development. Excitatory synaptic activity induces Npas4 in a Ca^{2+} -dependent manner, and the level of Npas4 determines the number of functional GABAergic synapses by controlling a program of activity-dependent gene expression. Future characterization of Npas4 target genes will help to determine whether Npas4 acts by initiating inhibitory synapse formation, stabilizing nascent inhibitory synapses or promoting the maturation of weak inhibitory synapses. It is possible that different subsets of Npas4 targets control the development of GABAergic synapses formed by distinct classes of interneurons, providing a mechanism for the independent regulation of inhibition received

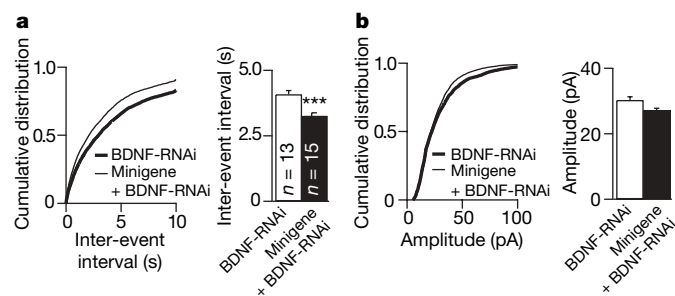


Figure 6 | Knockdown of BDNF partially attenuates the ability of the Npas4-minigene to elevate GABAergic synapses. **a**, Cumulative distributions (left) and mean \pm s.e.m. (right) of mIPSC inter-event intervals in neurons transfected with BDNF-RNAi (4086.1 \pm 140.7 ms) or BDNF-RNAi + Npas4-minigene (3257.9 \pm 117.7 ms). **b**, Cumulative distributions (left) and mean \pm s.e.m. (30.1 \pm 0.9 and 27.1 \pm 0.7 pA, BDNF-RNAi versus BDNF-RNAi + Npas4-minigene, right) of mIPSC amplitudes. Total numbers of neurons analysed in each condition (*n*) are indicated. ****P* < 0.001.

by subregions of a neuron. Another intriguing possibility is that Npas4 controls experience-dependent developmental processes, such as critical period plasticity in the visual cortex, which depend on the function and maturation of GABAergic synapses.

Although Npas4 is not required for the initial formation of excitatory synapses, activation of Npas4 in excitatory neurons within a neural circuit appears to diminish the excitatory synaptic input they receive. It is not known whether this effect is mediated directly by Npas4 or is an indirect consequence of changes in inhibitory input. In either case, our findings suggest that Npas4 functions as part of the homeostatic mechanism that stabilizes the activity of a neuron in the face of changing glutamatergic input. Further investigation of the function of Npas4 and other activity-dependent regulators of inhibitory synapses⁴⁵ will provide insight into the mechanism by which neuronal activity controls the balance between excitation and inhibition in the brain, and how the disruption of this balance leads to neurological disorders such as autism, epilepsy and schizophrenia.

METHODS SUMMARY

Dissociated neuron culture and transfection. Dissociated cortical and hippocampal neurons were prepared from E18 rat or E16 mouse embryos as previously described⁴⁶. Cultures were maintained in Neurobasal Medium supplemented with B27 (Invitrogen), penicillin–streptomycin and glutamine. Neurons were plated at 100,000–150,000 per well in a 24-well plate or 15,000,000 per 10-cm plate. For synapse assays, hippocampal neurons (60,000 per well, 24-well plate) were plated on a monolayer of astrocytes⁴⁷. Neurons were transfected at 5–6 days *in vitro* (DIV) using the calcium phosphate method⁴⁶.

Organotypic slice culture and transfection. Organotypic hippocampal slice cultures were prepared from P6 rats or mice as previously described⁴⁸. Slices were biolistically transfected with a Helios Gene Gun (Biorad) after 2 days. Bullets for the gene gun were 1.6- μ m gold particles coated with 15 μ g eGFP and either 5 or 10 μ g RNAi construct, 45 μ g Npas4-minigene or 30 μ g Cre. Empty plasmid was added to bring the total DNA to 60 μ g in each case.

Synapse density assay. Hippocampal neurons (14–18 DIV) were immunostained for synaptic markers and imaged on a Zeiss LSM5 Pascal microscope using a $\times 63$ objective lens. Image acquisition and synapse quantification were performed in a blinded manner. Glutamatergic neurons were identified based on morphology and the absence of cytosolic GAD65 staining. Synapse density was measured using Metamorph software as previously described⁴⁷. Within each experiment, density of puncta was normalized against the control, and the error of the control propagated into each experimental condition. Statistical analysis was performed on the raw data (without normalization) using multifactorial analysis of variance in StatView 4.5 (Abacus Concepts).

Electrophysiology. Whole-cell patch clamp recordings were made at room temperature from CA1 pyramidal neurons 7–10 days after transfection. Data were analysed in IgorPro (Wavemetrics) using custom-written macros. Statistical significance was determined by Kolmogorov–Smirnov test and Monte Carlo simulation.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 5 May; accepted 25 July 2008.

Published online 24 September 2008.

- Katz, L. C. & Shatz, C. J. Synaptic activity and the construction of cortical circuits. *Science* **274**, 1133–1138 (1996).
- Spitzer, N. C. Electrical activity in early neuronal development. *Nature* **444**, 707–712 (2006).
- Wong, R. O. & Ghosh, A. Activity-dependent regulation of dendritic growth and patterning. *Nature Rev. Neurosci.* **3**, 803–812 (2002).
- Zito, K. & Svoboda, K. Activity-dependent synaptogenesis in the adult mammalian cortex. *Neuron* **35**, 1015–1017 (2002).
- Ben-Ari, Y. Excitatory actions of GABA during development: the nature of the nurture. *Nature Rev. Neurosci.* **3**, 728–739 (2002).
- Kenet, T., Froemke, R. C., Schreiner, C. E., Pessah, I. N. & Merzenich, M. M. Perinatal exposure to a noncoplanar polychlorinated biphenyl alters tonotopy, receptive fields, and plasticity in rat primary auditory cortex. *Proc. Natl Acad. Sci. USA* **104**, 7646–7651 (2007).
- Maffei, A., Nelson, S. B. & Turrigiano, G. G. Selective reconfiguration of layer 4 visual cortical circuitry by visual deprivation. *Nature Neurosci.* **7**, 1353–1359 (2004).
- Buschges, A. & Manira, A. E. Sensory pathways and their modulation in the control of locomotion. *Curr. Opin. Neurobiol.* **8**, 733–739 (1998).
- Brown, P., Ridding, M. C., Werhahn, K. J., Rothwell, J. C. & Marsden, C. D. Abnormalities of the balance between inhibition and excitation in the motor cortex of patients with cortical myoclonus. *Brain* **119**, 309–317 (1996).
- Cline, H. Synaptogenesis: a balancing act between excitation and inhibition. *Curr. Biol.* **15**, R203–R205 (2005).
- Mohler, H. GABA_A receptors in central nervous system disease: anxiety, epilepsy, and insomnia. *J. Recept. Signal Transduct. Res.* **26**, 731–740 (2006).
- Rubenstein, J. L. & Merzenich, M. M. Model of autism: increased ratio of excitation/inhibition in key neural systems. *Genes Brain Behav.* **2**, 255–267 (2003).
- Wassef, A., Baker, J. & Kochan, L. D. GABA and schizophrenia: a review of basic science and clinical studies. *J. Clin. Psychopharmacol.* **23**, 601–640 (2003).
- Burrone, J., O'Byrne, M. & Murthy, V. N. Multiple forms of synaptic plasticity triggered by selective suppression of activity in individual neurons. *Nature* **420**, 414–418 (2002).
- Malinow, R. & Malenka, R. C. AMPA receptor trafficking and synaptic plasticity. *Annu. Rev. Neurosci.* **25**, 103–126 (2002).
- Turrigiano, G. G., Leslie, K. R., Desai, N. S., Rutherford, L. C. & Nelson, S. B. Activity-dependent scaling of quantal amplitude in neocortical neurons. *Nature* **391**, 892–896 (1998).
- Benevento, L. A., Bakum, B. W. & Cohen, R. S. Gamma-aminobutyric acid and somatostatin immunoreactivity in the visual cortex of normal and dark-reared rats. *Brain Res.* **689**, 172–182 (1995).
- Chattopadhyaya, B. *et al.* Experience and activity-dependent maturation of perisomatic GABAergic innervation in primary visual cortex during a postnatal critical period. *J. Neurosci.* **24**, 9598–9611 (2004).
- Foeller, E. & Feldman, D. E. Synaptic basis for developmental plasticity in somatosensory cortex. *Curr. Opin. Neurobiol.* **14**, 89–95 (2004).
- Marty, S., Wehrle, R. & Sotelo, C. Neuronal activity and brain-derived neurotrophic factor regulate the density of inhibitory synapses in organotypic slice cultures of postnatal hippocampus. *J. Neurosci.* **20**, 8087–8095 (2000).
- Micheva, K. D. & Beaulieu, C. Development and plasticity of the inhibitory neocortical circuitry with an emphasis on the rodent barrel field cortex: a review. *Can. J. Physiol. Pharmacol.* **75**, 470–478 (1997).
- Seil, F. J. & Drake-Baumann, R. Activity-dependent changes in “transplanted” cerebellar cultures. *Exp. Neurol.* **138**, 327–337 (1996).
- Hensch, T. K. Critical period plasticity in local cortical circuits. *Nature Rev. Neurosci.* **6**, 877–888 (2005).
- Turrigiano, G. G. & Nelson, S. B. Homeostatic plasticity in the developing nervous system. *Nature Rev. Neurosci.* **5**, 97–107 (2004).
- Benson, D. L. & Cohen, P. A. Activity-independent segregation of excitatory and inhibitory synaptic terminals in cultured hippocampal neurons. *J. Neurosci.* **16**, 6424–6432 (1996).
- Nedivi, E., Hevroni, D., Naot, D., Israeli, D. & Citri, Y. Numerous candidate plasticity-related genes revealed by differential cDNA cloning. *Nature* **363**, 718–722 (1993).
- Worley, P. F., Cole, A. J., Saffen, D. W. & Baraban, J. M. Regulation of immediate early genes in brain: role of NMDA receptor activation. *Prog. Brain Res.* **86**, 277–285 (1990).
- Ooe, N., Saito, K., Mikami, N., Nakatuka, I. & Kaneko, H. Identification of a novel basic helix-loop-helix-PAS factor, NXF, reveals a Sim2 competitive, positive regulatory role in dendritic-cytoskeleton modulator drebrin gene expression. *Mol. Cell. Biol.* **24**, 608–616 (2004).
- Flood, W. D., Moyer, R. W., Tsykin, A., Sutherland, G. R. & Koblar, S. A. Nxf and Fbxo33: novel seizure-responsive genes in mice. *Eur. J. Neurosci.* **20**, 1819–1826 (2004).

30. Hester, I. *et al.* Transient expression of Nxf, a bHLH-PAS transactivator induced by neuronal preconditioning, confers neuroprotection in cultured cells. *Brain Res.* **1135**, 1–11 (2007).
31. Shamloo, M. *et al.* Npas4, a novel helix-loop-helix PAS domain protein, is regulated in response to cerebral ischemia. *Eur. J. Neurosci.* **24**, 2705–2720 (2006).
32. Fletcher, T. L., Cameron, P., De Camilli, P. & Banker, G. The distribution of synapsin I and synaptophysin in hippocampal neurons developing in culture. *J. Neurosci.* **11**, 1617–1626 (1991).
33. Huang, Z. J., Di Cristo, G. & Ango, F. Development of GABA innervation in the cerebral and cerebellar cortices. *Nature Rev. Neurosci.* **8**, 673–686 (2007).
34. Gomeza, J. *et al.* Deletion of the mouse glycine transporter 2 results in a hyperekplexia phenotype and postnatal lethality. *Neuron* **40**, 797–806 (2003).
35. Kash, S. F. *et al.* Epilepsy in mice deficient in the 65-kDa isoform of glutamic acid decarboxylase. *Proc. Natl Acad. Sci. USA* **94**, 14060–14065 (1997).
36. Kash, S. F., Tecott, L. H., Hodge, C. & Baekkeskov, S. Increased anxiety and altered responses to anxiolytics in mice deficient in the 65-kDa isoform of glutamic acid decarboxylase. *Proc. Natl Acad. Sci. USA* **96**, 1698–1703 (1999).
37. Huang, Z. J. *et al.* BDNF regulates the maturation of inhibition and the critical period of plasticity in mouse visual cortex. *Cell* **98**, 739–755 (1999).
38. Rutherford, L. C., DeWan, A., Lauer, H. M. & Turrigiano, G. G. Brain-derived neurotrophic factor mediates the activity-dependent regulation of inhibition in neocortical cultures. *J. Neurosci.* **17**, 4527–4535 (1997).
39. Seil, F. J. & Drake-Baumann, R. TrkB receptor ligands promote activity-dependent inhibitory synaptogenesis. *J. Neurosci.* **20**, 5367–5373 (2000).
40. Aid, T., Kazantseva, A., Piirsoo, M., Palm, K. & Timmusk, T. Mouse and rat BDNF gene structure and expression revisited. *J. Neurosci. Res.* **85**, 525–535 (2007).
41. Liu, Q. R. *et al.* Rodent BDNF genes, novel promoters, novel splice variants, and regulation by cocaine. *Brain Res.* **1067**, 1–12 (2006).
42. Tao, X., Finkbeiner, S., Arnold, D. B., Shaywitz, A. J. & Greenberg, M. E. Ca^{2+} influx regulates BDNF transcription by a CREB family transcription factor-dependent mechanism. *Neuron* **20**, 709–726 (1998).
43. Zhou, P. *et al.* Polarized signaling endosomes coordinate BDNF-induced chemotaxis of cerebellar precursors. *Neuron* **55**, 53–68 (2007).
44. Kohara, K. *et al.* A local reduction in cortical GABAergic synapses after a loss of endogenous brain-derived neurotrophic factor, as revealed by single-cell gene knock-out method. *J. Neurosci.* **27**, 7234–7244 (2007).
45. Chubykin, A. A. *et al.* Activity-dependent validation of excitatory versus inhibitory synapses by neuroligin-1 versus neuroligin-2. *Neuron* **54**, 919–931 (2007).
46. Xia, Z., Dudek, H., Miranti, C. K. & Greenberg, M. E. Calcium influx via the NMDA receptor induces immediate early gene transcription by a MAP kinase/ERK-dependent mechanism. *J. Neurosci.* **16**, 5425–5436 (1996).
47. Paradis, S. *et al.* An RNAi-based approach identifies molecules required for glutamatergic and GABAergic synapse development. *Neuron* **53**, 217–232 (2007).
48. Stoppini, L., Buchs, P. A. & Muller, D. A simple method for organotypic cultures of nervous tissue. *J. Neurosci. Methods* **37**, 173–182 (1991).
49. Li, C. & Wong, W. H. Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. *Proc. Natl Acad. Sci. USA* **98**, 31–36 (2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank members of the Greenberg laboratory for suggestions; S. Paradis, J. M. Gray, S. S. Margolis, J. Zieg and C. M. Fletcher for reading the manuscript; S. Vasquez for preparing primary neuronal cell cultures; M. Thompson, Y. Zhou and H. Ye for assistance in generating Npas4^{-/-} mice; T. Diefenbach and the Neurobiology Program Imaging Center for assistance with confocal microscopy; M. Fagioli for help with dissection of the visual cortex; and X. J. Liu and C. Chen for help with electrophysiology. M.E.G. acknowledges the generous support of the F. M. Kirby Foundation to the Neurobiology Program of the Children's Hospital and support from the Nancy Lurie Marks Family Foundation. This work was supported by a Lefler Foundation postdoctoral fellowship (Y.L.), a Ruth L. Kirschstein National Research Service Award and a Helen Hay Whitney postdoctoral fellowship (B.L.B.), a National Science Foundation Graduate Research Fellowship (A.D.L.), the Jane Coffin Childs Memorial Fund (T.-K.K.) and Mental Retardation Research Center grant HD18655 and National Institutes of Health grants NS27572 and NS48276 (M.E.G.).

Author Contributions Y.L. and M.E.G. conceived and designed the experiments and wrote the manuscript. Y.L. performed or participated in each of the experiments described in the manuscript. B.L.B. performed the electrophysiological recordings and contributed to the writing of the manuscript. J.L.H. quantified Npas4 mRNA levels for the light stimulation experiment, generated the Npas4-minigene construct and performed the luciferase assay to characterize it, managed the Npas4 animal colony and provided extensive technical support. A.D.L. performed immunocytochemistry for the light stimulation experiment and confocal imaging of neurons in the synapse assay with Npas4-RNAi. A.C.K. provided technical support during the early phase of the study and helped generate many reagents used in this study including the Npas4 antibody, Npas4 knockout construct and Npas4-RNAi lentivirus. T.-K.K. performed the chromatin immunoprecipitation experiments. L.S.H. helped generate the Npas4 antibody. A.N.M. performed the initial chromatin immunoprecipitation experiments.

Author Information Data have been placed in the GEO database under accession numbers GSE11256 and GSE11258. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to M.E.G. (meg@hms.harvard.edu).

METHODS

DNA constructs. Npas4-RNAi, 5'-GGTTGACCCTGATAATTATA-3', and control-RNAi with a scrambled sequence (underlined), 5'-GGTTCAGCGTCATAA TTTA-3', were cloned into the pSuper expression vector (OligoEngine). The same Npas4-RNAi was used to generate the lentivirus construct (Cellogenetics, Inc). The control lentivirus construct, pLenti-shGL3 (Cellogenetics, Inc), targets the luciferase gene, which is not expressed by mouse hippocampal neurons.

The Npas4-minigene was generated by subcloning the mouse genomic region from CTCGGTTTCTATCTTCATGCC to GCCATGTGGCCTGCCGGTAC into a pBluescriptIIK+ vector between the *SacI* and *KpnI* digestion sites. The sequence targeted by Npas4-RNAi was mutated to AgtCgaTccGataattta, preserving the amino-acid sequence, to generate an RNAi-resistant Npas4-minigene.

The Npas4 luciferase reporter was constructed by replacing the tandem MEF2 responsive elements with three copies of the Npas4 responsive element²⁸ in the 3×MRE-luc plasmid⁵⁰.

Immunocytochemistry and immunohistochemistry. Cultured neurons were fixed and immunostained as previously described⁴⁷. For immunohistochemistry, mice were perfused intracardially with cold PBS followed by 4% paraformaldehyde in PBS. Brains were post-fixed in 4% paraformaldehyde in PBS at 4 °C overnight followed by incubation in 30% sucrose in PBS at 4 °C for 24 h. These cryo-protected brains were immediately sectioned on a cryostat (Leica) or stored in Tissue-Tec O.C.T. at -80 °C for later sectioning. Matching brain sections (35 µm) were incubated in blocking solution (3% BSA, 3% goat serum, 0.3% Triton X-100, 0.2% Tween-20 in PBS) at room temperature for 2 h. Subsequently, sections were incubated with primary antibodies overnight at 4 °C and with secondary antibodies at room temperature for 2 h, both in blocking solution. Finally, sections were mounted on slides with Aquamount (Lerner Laboratories).

The following antibodies were used: GAD65 (mouse, 1:1,000, Chemicon), GABA_A-receptor γ2 subunit (rabbit, 1:400, Chemicon), PSD95 (mouse, 1:300, Affinity BioReagents), synapsin I (rabbit, 1:500, Chemicon), c-fos (rabbit, 1:1,000, Santa Cruz sc-52), phosphor-serine 133 CREB (mouse, 1:2,000, Upstate). The Npas4 antibody (rabbit, 1:2,000) was raised against a carboxy-terminal region of Npas4 (amino acids 597–802) and its specificity was tested in Npas4^{-/-} mice (Supplementary Fig. 2).

Synapse density assay. Confocal image acquisition and synapse density measurement were performed as previously described⁴⁷. The perisomatic region of a neuron was defined as the cell body plus a 20-pixel circumference. Each pixel is approximately 0.143 µm. The perisomatic region was excluded from the excitatory synapse density analysis.

Because synapse density and immunostaining vary significantly between experiments, it is necessary to normalize each experiment before combining them. Normalization and error propagation were performed as previously described⁴⁷.

Electrophysiology. Slices were perfused with artificial cerebrospinal fluid containing (in mM) 127 NaCl, 25 NaHCO₃, 1.25 Na₂HPO₄, 2.5 KCl, 2 CaCl₂, 1 MgCl₂, 25 glucose, and saturated with 95% O₂, 5% CO₂. The internal solution for mIPSCs contained (in mM) 147 CsCl, 5 Na₂-phosphocreatine, 10 HEPES, 2 MgATP, 0.3 Na₂GTP and 1 EGTA. The internal solution for mEPSCs contained (in mM) 120 cesium methane sulfonate, 10 HEPES, 4 MgCl₂, 4 Na₂ATP, 0.4 Na₂GTP, 10 sodium phosphocreatine and 1 EGTA. Osmolarity and pH were adjusted to 300 mOsm and 7.3 with Millipore water and CsOH, respectively. mIPSCs were pharmacologically isolated by bath application of (in µM) 0.5 tetrodotoxin, 10 (R)-CPP and 10 NBQX disodium salt (all from Tocris Bioscience); mEPSCs were isolated with 0.5 tetrodotoxin and 50 picrotoxin

(Tocris Bioscience), and augmented with 10 cyclothiazide. Cells with series resistance larger than 25 MΩ during the recordings were discarded.

Data were analysed in IgorPro (Wavemetrics) using custom-written macros. For each trace, the event threshold was set at 1.5 times the root-mean-square current. Currents were counted as events if they crossed the event threshold, had a rapid rise time (1.5 pA ms⁻¹) and had an exponential decay (2 < τ < 200 ms, 1 < τ < 50 ms for mIPSC and mEPSC, respectively).

Statistical significance was determined by two methods. First, 50 random points selected from each cell were concatenated to describe the cumulative distributions of events in each condition and then compared by a Kolmogorov–Smirnov test. Second, a Monte Carlo simulation was performed in which points were randomly sampled from each condition and the mean of these samples compared at least 1,000 times. *P* < 0.05 from both tests was considered significant.

DNA microarrays and data analysis. For all microarray experiments, total RNA was purified using RNeasy mini kits (Qiagen) and biotin-labelled cRNA was generated following Affymetrix standard protocols. Ten micrograms of the labelled cRNA was hybridized to Affymetrix mouse MOE430 arrays.

To identify activity-regulated genes, P0 mouse cortical neurons were cultured for 7 days and then depolarized with 50 mM KCl for 1 and 6 h. To identify Npas4-regulated genes, E16 mouse hippocampal neurons were infected (3 DIV) with Npas4-RNAi or control lentivirus, depolarized (8 DIV) with 50 mM KCl, and total RNA was collected 0, 1, 3 and 6 h later. Drebrin, an Npas4 target previously identified in a neuroblastoma cell line²⁸ was not affected by Npas4-RNAi in post-mitotic neurons.

The DNA-Chip (dChip) software package⁴⁹ was used to analyse the microarray data. Genes were considered candidates if they met the following criteria: (1) there was at least a 1.5-fold difference between the experimental and control conditions (depolarized versus non-depolarized, or Npas4-RNAi virus versus control virus) at any of the time points; and (2) there was an absolute difference of more than 100 normalized hybridization intensity units in the expression level between the experimental and control samples.

Chromatin immunoprecipitation. Rat cortical neurons were treated with 1 µM tetrodotoxin and 100 µM AP5 overnight and then stimulated with 55 mM KCl for 2 h at 8 DIV. Chromatin immunoprecipitation was performed as previously described⁴⁹. The following primers were used for quantitative PCR: Chr3, forward: 5'-GACCCCATCCTGTGGTTATG-3', reverse: 5'-GCAACAAAGGCA AATGGAAT-3'; Npas4, forward: 5'-CAGGATGACTCACACTGACAGTATT TTAG-3', reverse: 5'-GTGGGAGAAGAGCTATTTATATCACCAG-3'; BDNF promoter I, forward: 5'-GTGCCTCTCGCCTAGTCATC-3', reverse: 5'-AGGG AACAACTGCGTGAATC-3'; BDNF promoter IV, forward: 5'-CAGGGAGC TACTCACCAAC-3', reverse: 5'-GCACACAGAAGCCAAACCTT-3'; BDNF coding region, forward: 5'-GACAAGGCAACTTGGCCTAC-3', reverse: 5'-TC GTCAGACCTCTCGAACCT-3'; BDNF 3' untranslated region, forward: 5'-GA AACGCCACACCTACGAAT-3', reverse: 5'-GGCACC GAATCCAATTCTAA-3'.

Quantitative reverse transcriptase PCR. The mRNA level of the gene of interest was normalized against the mRNA level of β-tubulin in the same sample. The following primers were used: Npas4, forward: 5'-GCTATA CTCAGAAGGTCCAGAAGGC-3', reverse: 5'-TCAGAGAATGAGGGTAGCA CAGC-3'; BDNF, forward: 5'-GATGCCGCAACATGTCTATGA-3', reverse: 5'-TAATACTGTACACACGCTCAGCTC-3'; β-tubulin, forward: 5'-CGAC AATGAAGCCCTCTACGAC-3', reverse: 5'-ATGGTGGCAGACACAAGGTGG TTG-3'.

50. Flavell, S. W. *et al.* Activity-dependent regulation of MEF2 transcription factors suppresses excitatory synapse number. *Science* **311**, 1008–1012 (2006).

Glycogen synthase kinase 3 in *MLL* leukaemia maintenance and targeted therapy

Zhong Wang¹, Kevin S. Smith¹, Mark Murphy¹, Obdulio Piloto¹, Tim C. P. Somervaille¹ & Michael L. Cleary¹

Glycogen synthase kinase 3 (GSK3) is a multifunctional serine/threonine kinase that participates in numerous signalling pathways involved in diverse physiological processes. Several of these pathways are implicated in disease pathogenesis, which has prompted efforts to develop GSK3-specific inhibitors for therapeutic applications. However, before now, there has been no strong rationale for targeting GSK3 in malignancies. Here we report pharmacological, physiological and genetic studies that demonstrate an oncogenic requirement for GSK3 in the maintenance of a specific subtype of poor prognosis human leukaemia, genetically defined by mutations of the *MLL* proto-oncogene. In contrast to its previously characterized roles in suppression of neoplasia-associated signalling pathways, GSK3 paradoxically supports *MLL* leukaemia cell proliferation and transformation by a mechanism that ultimately involves destabilization of the cyclin-dependent kinase inhibitor p27^{Kip1}. Inhibition of GSK3 in a preclinical murine model of *MLL* leukaemia provides promising evidence of efficacy and earmarks GSK3 as a candidate cancer drug target.

GSK3 is a serine/threonine kinase that functions in numerous signalling pathways initiated by diverse stimuli¹. Originally studied for its role in glycogen metabolism and insulin action, GSK3 has subsequently been shown to have central functions in many cellular and physiological processes including transcription, cell cycle division, apoptosis, cell fate determination and stem cell maintenance, among others^{1–3}. GSK3 is constitutively active in resting cells, showing a preference for primed substrates⁴, and is functionally inactivated after phosphorylation by various kinases in response to different stimuli^{3,5}. Given its various contributions and the diversity of putative substrates, many levels of regulation help confer GSK3 signalling specificity, which varies among cell types and their states of differentiation.

GSK3 functions in several pathways implicated in human diseases, which has prompted efforts to develop specific inhibitors for therapeutic applications. GSK3 facilitates non-insulin-dependent diabetes by the inactivation of glycogen synthase^{3,6}, and may have a role in promoting various inflammatory processes through the activation of the transcription factor nuclear factor- κ B by, at present, undefined mechanisms^{7,8}. GSK3-mediated hyperphosphorylation of tau (also known as MAPT), a component of neurofibrillary tangles, may facilitate Alzheimer's disease and other neurodegenerative disorders⁹. In cancer cells, however, signalling pathways that are normally suppressed by GSK3—such as Wnt and Hedgehog, which are involved in embryonic cell fate determination and normal stem cell maintenance—are aberrantly activated^{10–13}. This underscores the normal role of GSK3 in mediating phosphorylation of substrates such as β -catenin (Wnt signalling), MYCN (Hedgehog signalling) and JUN, which leads to their destruction and/or inactivation, thus inhibiting signals that otherwise promote proliferation and self-renewal^{14–16} (Supplementary Fig. 1). Consistent with these molecular functions, GSK3 inhibition significantly enhances maintenance of embryonic stem cell pluripotency and haematopoietic stem cell repopulation after bone marrow transplantation^{17,18}, although the specific pathways for these effects remain undefined. Despite its inhibitory roles in

pathways implicated in cancer pathogenesis, there has so far been no compelling rationale for the targeting of GSK3 as a therapeutic approach in malignancies. Here we demonstrate a paradoxical and unexpected role for GSK3 in cancer maintenance, and we establish GSK3 as a potential selective therapeutic target in a genetically distinctive and poor prognosis subset of acute leukaemia.

GSK3 inhibition induces G1 arrest of *MLL* leukaemia cells

A small-scale screen was conducted to identify compounds that specifically blocked the growth of genetically defined subsets of leukaemia cells. Thirty compounds (Supplementary Table 1) that target principal kinases or other enzymes were screened for differential dose-responses in various cell lines (Supplementary Table 2). These cell lines represent human leukaemias harbouring a variety of chromosomal translocations that create distinctive chimaeric fusion proteins implicated in disease pathogenesis. The leukaemia cell lines were comparably sensitive to most of the tested compounds (data not shown). However, cell lines that expressed *MLL*-AF4 or *MLL*-AF5, the highly related fusion oncogenes created by t(4;11) or t(5;11) chromosomal translocations, respectively, showed enhanced sensitivity to GSK3-IX, a GSK3 inhibitor that also targets cyclin-dependent kinases (CDKs; Fig. 1a; for clarity, only two representative control cell lines are shown). Their proliferation was inhibited at a half-maximal inhibitory concentration (IC₅₀) of 0.3–2 μ M, a concentration range comparable to that which promotes expansion of haematopoietic stem cells *in vitro*¹⁸, but tenfold lower than the toxicity levels for non-*MLL* leukaemia cell lines (Fig. 1a) and normal bone marrow progenitors (see later). In contrast, the CDK inhibitors roscovitine (Fig. 1a), flavopiridol and olomoucine (data not shown) had similar IC₅₀ values for all cell lines, suggesting that the inhibitory effects of GSK3-IX on *MLL* cell lines resulted from GSK3, not CDK, blockade. Further studies with SB216763 (a widely used maleimide-containing GSK3 inhibitor with a relatively higher IC₅₀ than GSK3-IX) and with alsterpaullone (which has a similar inhibition profile as GSK3-IX)

¹Department of Pathology, Stanford University School of Medicine, Stanford, California 94305, USA.

confirmed that *MLL* leukaemia cells were differentially sensitive to GSK3 inhibition (Fig. 1b and data not shown). Increased β -catenin levels correlated with effective GSK3 inhibition, which did not alter *MLL* oncoprotein abundance or function (Supplementary Fig. 2a, b).

Cell cycle analyses showed a marked reduction in G1–S phase progression of *MLL* leukaemia cells after 24 h of inhibitor treatment, whereas non-*MLL* leukaemia cells were only minimally affected (Fig. 1c). More prolonged incubation with inhibitor (6 days) was associated with cell death, as evidenced by a substantial increase in sub-G0/G1 DNA content (Supplementary Fig. 3a, b). These data suggest that GSK3, which is constitutively active in normal resting cells, paradoxically supports the proliferation and sustained survival of a genetically defined subset of leukaemia.

GSK3 dependence is a general feature of *MLL* transformed cells

Murine transformation models were used to characterize the role of GSK3 in *MLL* leukaemia further. Transduction of *MLL* oncogenes

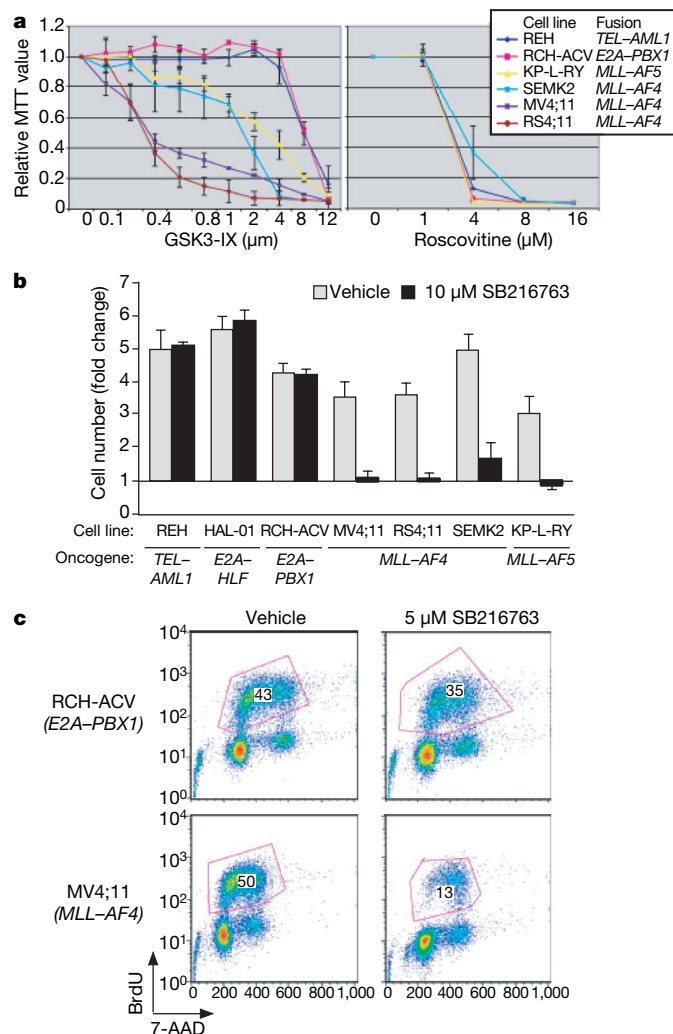


Figure 1 | Sensitivity of *MLL* leukaemia cell lines to GSK3 inhibition. **a**, The growth of human leukaemia cell lines was assessed after 3 days culture in the presence of the indicated concentrations of GSK3-IX (left panel) or roscovitine (right panel). The results are expressed as the cell numbers relative to those without drug treatment, and represent the mean of three independent experiments (\pm s.e.m.). MTT, 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide. **b**, Human leukaemia cell lines were cultured in the absence or presence of 10 μ M SB216763 for 2 days. The results of a representative experiment are expressed as the fold change in cell number compared to day 0 (\pm s.e.m., $n = 3$). **c**, Human leukaemia cell lines were cultured in the presence of 5 μ M SB216763 for 24 h, and BrdU incorporation was quantified by fluorescence-activated cell sorting (FACS) analysis.

1206

into primary murine myeloid progenitors induces aberrant *Hox* gene expression^{19,20}, leading to enhanced self-renewal *in vitro* and acute myeloid leukaemias *in vivo* that accurately model the features of human *MLL* leukaemia²¹ without altering GSK3 levels or activity (Supplementary Fig. 3c, d). Culture of *MLL*-transduced myeloid progenitors with a GSK3 inhibitor reduced their clonogenic potentials and proliferation (Fig. 2a). This contrasted with progenitors immortalized by other fusion oncogenes (Fig. 2a and data not shown), which showed no adverse growth effects with 10 μ M SB216763 treatment, as was also the case for primary myeloid progenitors (Supplementary Fig. 4). Inhibition of GSK3 primarily resulted in proliferative arrest of *MLL* transformed cells, but prolonged exposure induced morphological features of myeloid differentiation (Fig. 2b, c) and reduced expression of c-Kit (also known as KIT) (data not shown), a phenotypic marker of normal progenitors and *MLL* leukaemia stem cells²². Mouse B cell progenitors transformed by *MLL*-AF4, but not by other oncogenes, also showed markedly reduced proliferation in 10 μ M SB216763 (Fig. 2d). These data suggest that GSK3 dependence may be a primary consequence and general feature of *MLL* transformation in several haematopoietic lineages. Furthermore, expression of a constitutively active mutant of the protein kinase AKT, which phosphorylates GSK3 and negatively regulates its kinase activity, resulted in suppression of cell growth and clonogenic potentials of mouse myeloid and B cell progenitors transformed by *MLL* oncogenes (Supplementary Fig. 5), providing support that *MLL* transformed cells are dependent on GSK3 for continued proliferation and maintenance of their transformed phenotypes *in vitro*.

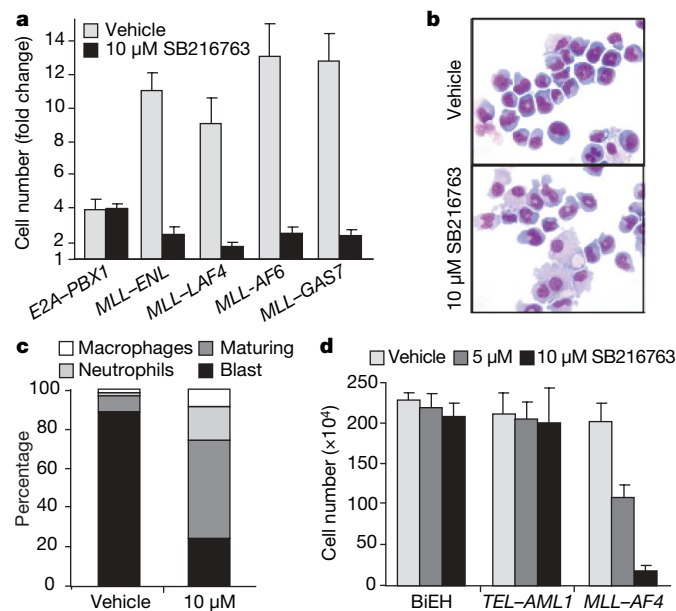


Figure 2 | Sensitivity of *MLL*-transformed mouse B and myeloid progenitors to GSK3 inhibition. **a**, The growth of myeloid progenitors transformed by various oncogenes was assessed after 3 days of culture in the presence or absence of a GSK3 inhibitor. The results of a representative experiment are expressed as the fold change in cell number compared to day 0 (\pm s.e.m., $n = 3$). **b**, **c**, The morphological features of *MLL*-ENL transformed myeloid progenitors were assessed after 4 days of culture in the presence or absence of GSK3 inhibitor. Original magnification, $\times 40$. The bar graph (c) indicates the mean number of cells with the indicated morphological features ($n = 3$). **d**, The growth of B lymphoid progenitors transformed by E2A-HLF and BCL2 (BiEH)³⁷, TEL-AML1 or MLL-AF4 oncogenes was assessed after 3 days of culture in the absence or presence (5 μ M or 10 μ M) of SB216763. The results of a representative experiment are expressed as the fold change in cell number compared to day 0 (\pm s.e.m., $n = 3$).

GSK3 α/β isoforms cooperatively maintain *MLL* transformation

To investigate further the GSK3 requirement, myeloid progenitors were isolated from fetal livers of *Gsk3b*^{-/-} mice (embryonic day (E)16 embryos), transduced with retroviral vectors encoding *MLL* or unrelated oncogenes (Fig. 3a), and then serially replated in methylcellulose culture to assess their self-renewal properties. *Gsk3b*^{-/-} cells were capable of sustaining the enhanced self-renewal typically induced by *MLL* oncogenes, and did not show reduced clonogenic potentials compared with wild-type cells transduced with the same *MLL* oncogenes (Supplementary Fig. 6) despite a 50% reduction in overall GSK3 activity levels (Supplementary Fig. 7d). Thus, GSK3- β was not required to initiate *MLL* transformation *in vitro*. However, *MLL*-transformed *Gsk3b*^{-/-} cells showed markedly increased sensitivity to pharmacological GSK3 inhibition (Fig. 3b, c), which was reversed by the forced expression of exogenous GSK3- β (Fig. 3d). In contrast, *Gsk3b*^{-/-} cells transformed by other fusion oncogenes (Fig. 3b and data not shown) were unaffected by a several fold higher concentration of inhibitor. Thus, genetic reduction of GSK3- β levels, by knockout or short-hairpin-RNA-mediated knockdown (Supplementary Fig. 7a), in *MLL*-transformed myeloid progenitors resulted in increased sensitivity to pharmacological GSK3 inhibition.

Persistence of the transformed phenotype but enhanced inhibitor sensitivity in the absence of GSK3- β suggested that the two GSK3 isoforms probably have redundant roles in *MLL* transformation. Thus, GSK3- α knockdown studies were performed in myeloid progenitors, which resulted in efficient reduction of GSK3- α protein levels in wild-type as well as *Gsk3b*^{-/-} cells, (Fig. 4a) accompanied by further decrease in total GSK3 activity to less than 20% of wild-type cells (Supplementary Fig. 7d). Unlike GSK3- β knockout or knockdown cells, *MLL*-transformed cells deficient for GSK3- α (*Gsk3a*^{KD}) did not show differences in growth or heightened sensitivity

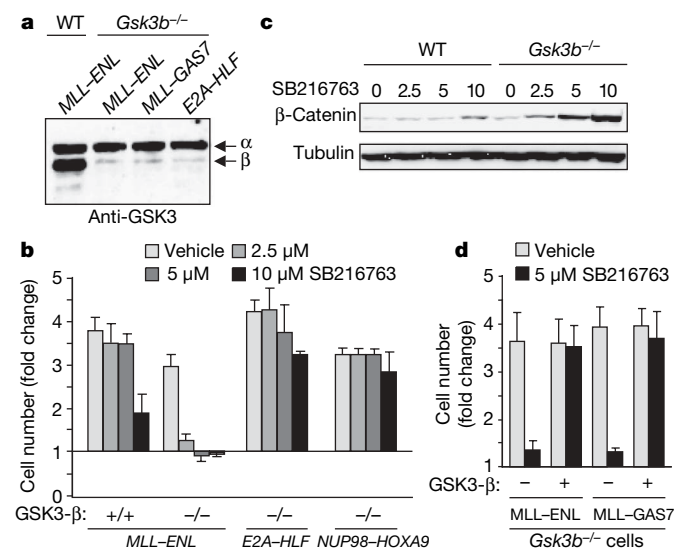


Figure 3 | Genetic ablation of GSK3- β hypersensitizes *MLL*-transformed cells to pharmacological GSK3 inhibition. **a**, Western blot analysis demonstrates the amounts of GSK3 protein isoforms in wild type or *Gsk3b*^{-/-} myeloid progenitors transformed by the indicated oncogenes. **b**, Wild type (+/+) or *Gsk3b*^{-/-} myeloid progenitors transformed by various oncogenes were incubated in the presence of the indicated concentrations of SB216763. Cell numbers were enumerated on day 2 and expressed as the fold change compared to day 0 (\pm s.e.m. of triplicate analyses). **c**, Western blot analysis demonstrates the relative amounts of β -catenin after treatment with the indicated concentrations (μ M) of inhibitor in wild type (WT) or *Gsk3b*^{-/-} myeloid progenitors transformed by *MLL*-ENL. **d**, *Gsk3b*^{-/-} myeloid progenitors transformed by *MLL* oncogenes were stably transduced with Flag-GSK3- β (+) or vector (-), and then incubated in the presence or absence of 5 μ M SB216763. Cell numbers were enumerated on day 2 and expressed as the fold change compared to day 0 (\pm s.e.m. of triplicate analyses).

to GSK3 inhibitors (Supplementary Fig. 7a, b). However, myeloid progenitors deficient for both GSK3 isoforms (*Gsk3b*^{-/-} *Gsk3a*^{KD}) showed a marked impairment in clonogenicity and proliferation compared to wild-type cells transformed by *MLL* oncogenes (Fig. 4b, c), and were unable to sustain long-term growth in culture. In contrast, the growth of cells transformed by other leukaemia oncogenes was unaffected by the compound deficiency of both GSK3 isoforms (Fig. 4b, c) despite substantially decreased GSK3 activity (Supplementary Fig. 7d). *Gsk3b*^{-/-} *Gsk3a*^{KD} cells transduced with *MLL* oncogenes also had a more differentiated myeloid morphology (Fig. 4d) and phenotype (not shown). Notably, *MLL*-ENL transformed cells lacking both GSK3 isoforms were unable to induce leukaemia in transplanted mice (Fig. 4e). Thus, GSK3 isoforms cooperatively maintain critical features of the *MLL* transformed phenotype, although GSK3- β serves a predominant role.

Efficacy of GSK3 inhibition in a mouse model of *MLL* leukaemia

A mouse model of *MLL*-associated leukaemia was used to assess whether GSK3 inhibition *in vivo* would affect the course of disease.

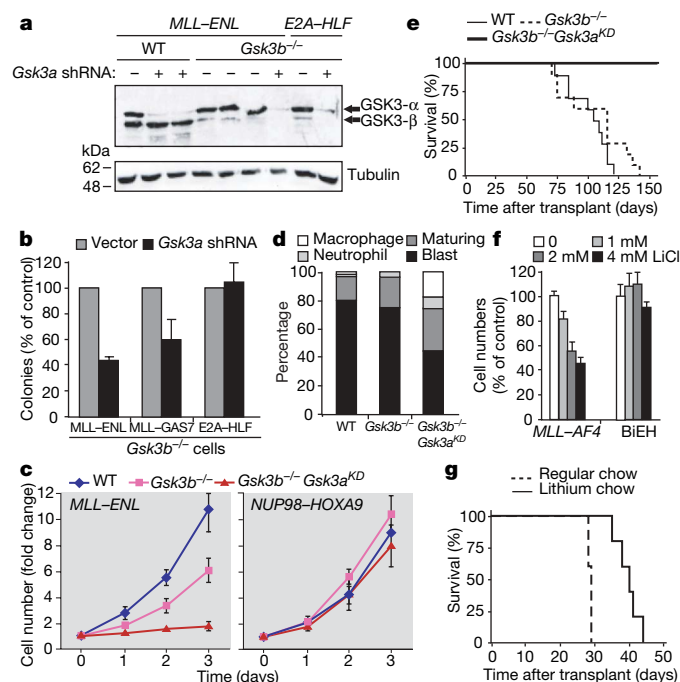


Figure 4 | Compound genetic deficiency of GSK3- α and GSK3- β impairs the growth and leukemogenicity of *MLL*-transformed cells. **a**, Western blot analysis was performed on wild type (WT) and *Gsk3b*^{-/-} myeloid progenitors transformed by the indicated oncogenes (top) and transduced by lentiviral vectors lacking (-) or expressing (+) *Gsk3a* shRNA. **b**, Myeloid progenitors transformed by the indicated oncogenes were acutely transduced with lentiviral vectors lacking or expressing *Gsk3a* shRNA and then plated in methylcellulose medium. Colonies were enumerated after 5 days, and the mean (\pm s.e.m.) numbers of three independent determinations are expressed relative to vector alone. **c**, Proliferation of myeloid progenitors (WT, *Gsk3b*^{-/-} or *Gsk3b*^{-/-} *Gsk3a*^{KD}) transformed by *MLL*-ENL (left panel) or NUP98-HOXA9 (right panel) was assessed at the indicated days in liquid culture (\pm s.e.m. of triplicate analyses). **d**, The morphological features of *MLL*-ENL transformed myeloid progenitors with the indicated genotypes were assessed after 4 days of culture. The bar graph indicates the mean number of cells with the indicated morphological features ($n = 3$). **e**, Survival curves are shown for cohorts of mice transplanted with cells (WT, *Gsk3b*^{-/-} or *Gsk3b*^{-/-} *Gsk3a*^{KD}) stably transduced with *MLL*-ENL (10 mice each). **f**, Cell numbers were determined after 3 days of culture in the indicated concentrations of lithium chloride. **g**, Survival curves show significantly different latencies ($P < 0.001$) for the development of acute leukaemia in cohorts of mice transplanted with *MLL*-AF4 leukaemia cells (5×10^4) and maintained on normal or lithium carbonate (0.4%) laced chow as indicated.

Mice transplanted with *MLL*-*AF4*-transformed B cell precursors developed a fatal aggressive leukaemia within 29 days, characterized by massive infiltration of the bone marrow, spleen and liver, with leukaemic blasts (data not shown). However, treatment with lithium carbonate, which has been extensively used to modulate *in vivo* GSK3 kinase activity²³ and impairs *MLL* leukaemia cell proliferation *in vitro* (Fig. 4f), resulted in a significant prolongation of survival (Fig. 4g and Supplementary Fig. 7e). These results indicate that sensitivity of *MLL*-transformed cells to GSK3 inhibition is not restricted to *in vitro* environments, and provide evidence of therapeutic efficacy.

p27^{Kip1} mediates the response to GSK3 inhibition

Cell cycle arrest in response to GSK3 inhibition suggested that cell cycle regulators may be downstream targets of GSK3 signalling in *MLL*-transformed cells. Western blot analysis implicated the CDK inhibitor (CDKI) p27^{Kip1} as the levels of this significantly increased in human *MLL* leukaemia cells (Fig. 5a) and murine transformed progenitors (Fig. 5b) after inhibitor treatment, which is temporally consistent with the onset of cell cycle arrest (Fig. 1c and Supplementary Fig. 3a). Conversely, p27^{Kip1} levels did not increase in control cells (Fig. 5a, b), which continued to actively cycle in the presence of inhibitor (Supplementary Fig. 3 and data not shown). GSK3, either directly or indirectly, negatively regulates p27^{Kip1} protein stability because inhibitor treatment increased the p27^{Kip1} half-life without inducing increased messenger RNA levels (Supplementary Fig. 8). β -Catenin levels increased substantially in *MLL*-transformed but also in control cells in response to GSK3 inhibition, whereas levels of p21, another CDKI, were not altered (Fig. 5a, b). Knockdown of p27^{Kip1} resulted in substantial reductions of p27^{Kip1} protein (Fig. 5c), and prevented the growth arrest otherwise induced by GSK3 inhibitor in *MLL*-transformed myeloid progenitors

(Fig. 5d). Thus, p27^{Kip1} is a critical downstream mediator of the cell cycle arrest associated with GSK3 inhibition in *MLL* transformed cells.

Discussion

GSK3 maintenance of a genetically distinctive subset of acute leukaemia establishes an enabling role for this multifunctional kinase in oncogenesis. This contrasts with its well-characterized function to suppress signalling pathways that otherwise promote proliferation and self-renewal, a role thwarted in colon cancer and other cancers with oncogenic mutations of β -catenin that abrogate its GSK3-mediated phosphorylation and subsequent destruction on the Wnt pathway²⁴. Similarly, hyperactivation of AKT is implicated in cancer pathogenesis through enhanced survival and proliferation²⁵. In contrast, *MLL*-transformed cells are sustained by GSK3 and consistently antagonized by GSK3 inhibitors of varying selectivity and specificity (Supplementary Fig. 1), and also adversely affected by constitutively active AKT, a physiological inhibitor of GSK3 activity. Genetic and pharmacological studies confirm the requirement of GSK3 to maintain *MLL*-mediated transformation and leukaemogenesis in preclinical murine models. Thus, GSK3 can promote oncogenesis and does not have an exclusively suppressive role in cancer pathogenesis.

The mechanism by which GSK3 supports *MLL*-oncogene-induced proliferation and transformation is mediated through the destabilization of p27^{Kip1}, a CDKI with established roles in tumour suppression²⁶. The p27^{Kip1} and p18^{Ink4c} genes have been shown to be direct transcriptional targets of *MLL*²⁷, a histone methyltransferase that positively maintains gene expression through covalent chromatin modification²⁰. In endocrine neoplasias, this tumour suppressor circuit that normally maintains CDKI expression is abrogated by mutations or loss of menin (also known as multiple endocrine neoplasia)²⁷, a critical component of the *MLL* histone methyltransferase complex^{28,29}. *MLL* oncoproteins also activate p27^{Kip1} expression³⁰, which would seem to be counterproductive for leukaemia pathogenesis. Our results indicate a potential mechanism for the suppression of p27^{Kip1} either directly or indirectly by GSK3, which provides a permissive cellular context for *MLL*-mediated transformation. However, phosphorylation of p27^{Kip1} by GSK3 has recently been shown to enhance its stability and prohibit cell cycle progression in the absence of growth factors³¹, which contrasts with increased p27^{Kip1} levels after GSK3 inhibition in *MLL*-transformed cells. Paradoxically increased p27^{Kip1} expression has also been observed in a myeloma cell line on GSK3 inhibition as part of a paracrine/autocrine feedback loop involving IL-6 signalling and forkhead transcription factors³². Therefore, the functional relationships of GSK3 with p27^{Kip1} seem complex and cell context dependent. Nevertheless, our studies link these factors on a pathogenic pathway that is critical for maintenance of *MLL* leukaemia.

The observed dependence on GSK3 provides a potential therapeutic target in a genetically distinctive subset of leukaemia defined by mutations of the *MLL* proto-oncogene. *MLL* is activated by a substantial array of chromosomal aberrations in diverse haematological disorders that account for approximately 5%–10% of sporadic leukaemias in adults and children³³. Independent of their association with other high-risk features, *MLL* aberrations are often predictive of poor clinical outcome³⁴, which warrants a search for new treatment approaches. GSK3 has not previously been considered as a therapeutic target in cancer. In fact, its normally suppressive roles in Wnt, hedgehog and Notch pathway signalling have raised the theoretical possibility that GSK3 inhibition may increase the risk of neoplasia. However, chronic administration of lithium, a relatively nonspecific GSK3 inhibitor used for the treatment of bipolar disorders, has not been associated with increased cancer risk³⁵. Notably, GSK3 is a specific *in vivo* modulator of haematopoietic stem cell activity, and GSK3 inhibitors enhance haematopoietic stem cell repopulation in mice after bone marrow transplantation¹⁷. Thus, like PTEN³⁶, GSK3 has converse roles in normal versus leukaemia stem cell maintenance, which may confer significant therapeutic

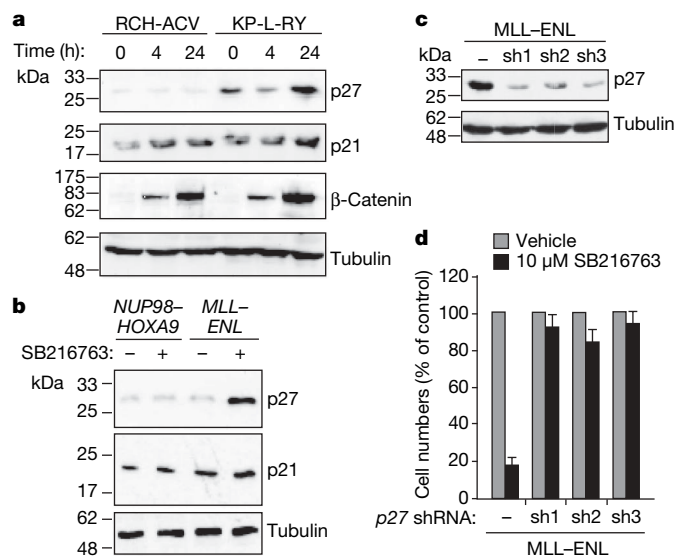


Figure 5 | GSK3 maintains *MLL* transformation through suppression of p27^{Kip1}. **a**, Human leukaemia cell lines (RCH-ACV and KP-L-RY) were treated with 10 μ M SB216763 for the indicated times and protein levels were assessed by western blot analysis. **b**, Murine myeloid progenitors transformed by the indicated oncogenes were cultured in the presence (+) or absence (–) of SB216763 (10 μ M) for 24 h in liquid culture, and then subjected to western blot analysis. **c**, Western blot analysis demonstrates the p27^{Kip1} (upper panel) or β -tubulin (lower panel, loading control) protein levels in *MLL*-*ENL* transformed myeloid progenitors stably transduced with lentiviral vectors expressing shRNAs specific for p27^{Kip1} (denoted as sh1 to sh3). **d**, Myeloid progenitors transformed by *MLL*-*ENL* and stably transduced with lentiviral vectors lacking (–) or expressing one of three different p27^{Kip1} shRNAs were cultured for 3 days in the presence or absence of 10 μ M SB216763. Viable cell numbers are expressed relative to the numbers obtained with lentiviral vector transduced cells (\pm s.e.m. of triplicate analyses).

selectivity. Our preclinical studies using lithium carbonate to target GSK3 in a murine model of *MLL* leukaemia provide promising evidence of efficacy. The paradoxical sensitivity of *MLL* leukaemias to GSK3 inhibition earmarks this multifunctional kinase as a therapeutic target, and provides a rationale to develop inhibitors with suitable pharmacodynamic properties for clinical trials to determine whether GSK3 inhibition may have therapeutic efficacy in a distinctive subset of poor prognosis leukaemia.

METHODS SUMMARY

Cell cultures and inhibitor assays. Human leukaemia cell lines, and transformed mouse myeloid or B cell precursors, were cultured in medium (with appropriate supplements) containing kinase inhibitors at the indicated concentrations. Cell viabilities were determined by trypan-blue dye exclusion and cell growth was quantified using MTT (3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyl-tetrazolium bromide) assays. Cell proliferation was determined by measuring BrdU incorporation. Flow cytometry was used to assess cell cycle status on the basis of propidium iodide staining and to quantify apoptosis on the basis of annexin V staining²⁷.

Transformation and leukaemogenesis assays. Myeloid progenitors were transduced with retroviral vectors as described previously²¹ with minor modifications, and were cultured in liquid or semi-solid medium supplemented with cytokines. B cell progenitors were transduced as described previously³⁷ and co-cultured on neo-resistant-irradiated OP9 stromal cells. After continuous passage and adaptation to liquid culture, immortalized cells were used for injections of syngeneic mice, and cell lines generated by explantation of splenocytes collected from leukaemic mice were used for GSK3 inhibitor studies. For knockdown studies, transformed progenitors were transduced with shRNA lentiviral constructs, selected for drug resistance *in vitro* and then evaluated for growth in the presence of GSK3 inhibitors. For *in vivo* studies, myeloid progenitors (wild type or *Gsk3b*^{-/-}) transformed by *MLL-ENL* were transduced with lentiviral knock-down constructs, selected for drug resistance and then transplanted (10⁶ cells) by intravenous injection into sub-lethally irradiated severe combined immunodeficient (SCID) mice. For lithium treatment, irradiated mice were transplanted with *MLL-AF4* leukaemic B cell progenitors (50,000 cells) and maintained on 0.4% lithium-carbonate-containing chow with saline water.

Protein assays. Protein extracts were prepared by cell lysis in buffer containing protease inhibitors, subjected to SDS-PAGE and analysed by western blot using primary antibodies as indicated throughout.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 9 April; accepted 18 July 2008.

Published online 17 September 2008.

- Doble, B. W. & Woodgett, J. R. GSK-3: tricks of the trade for a multi-tasking kinase. *J. Cell Sci.* **116**, 1175–1186 (2003).
- Kim, L. & Kimmel, A. R. GSK3 at the edge: regulation of developmental specification and cell polarization. *Curr. Drug Targets* **7**, 1411–1419 (2006).
- Cohen, P. & Frame, S. The renaissance of GSK3. *Nature Rev. Mol. Cell Biol.* **2**, 769–776 (2001).
- Fiol, C. J., Mahrenholz, A. M., Wang, Y., Roeske, R. W. & Roach, P. J. Formation of protein kinase recognition sites by covalent modification of the substrate. Molecular mechanism for the synergistic action of casein kinase II and glycogen synthase kinase 3. *J. Biol. Chem.* **262**, 14042–14048 (1987).
- Cross, D. A., Alessi, D. R., Cohen, P., Andjelkovich, M. & Hemmings, B. A. Inhibition of glycogen synthase kinase-3 by insulin mediated by protein kinase B. *Nature* **378**, 785–789 (1995).
- Kaidanovich, O. & Eldar-Finkelman, H. The role of glycogen synthase kinase-3 in insulin resistance and type 2 diabetes. *Expert Opin. Ther. Targets* **6**, 555–561 (2002).
- Hoeflich, K. P. *et al.* Requirement for glycogen synthase kinase-3 β in cell survival and NF- κ B activation. *Nature* **406**, 86–90 (2000).
- Martin, M., Rehani, K., Jope, R. S. & Michalek, S. M. Toll-like receptor-mediated cytokine production is differentially regulated by glycogen synthase kinase 3. *Nature Immunol.* **6**, 777–784 (2005).
- De Ferrari, G. V. & Inestrosa, N. C. Wnt signaling function in Alzheimer's disease. *Brain Res. Rev.* **33**, 1–12 (2000).
- Miller, J. R. & Moon, R. T. Signal transduction through β -catenin and specification of cell fate during embryogenesis. *Genes Dev.* **10**, 2527–2539 (1996).
- Jia, J. *et al.* Shaggy/GSK3 antagonizes Hedgehog signalling by regulating Cubitus interruptus. *Nature* **416**, 548–552 (2002).
- Miller, J. R. The Wnts. *Genome Biol.* **3**, 3001 (2002).
- Yost, C. *et al.* The axis-inducing activity, stability, and subcellular distribution of β -catenin is regulated in *Xenopus* embryos by glycogen synthase kinase 3. *Genes Dev.* **10**, 1443–1454 (1996).
- van Noort, M., Meeldijk, J., van der Zee, R., Destree, O. & Clevers, H. Wnt signaling controls the phosphorylation status of β -catenin. *J. Biol. Chem.* **277**, 17901–17905 (2002).
- Sears, R. *et al.* Multiple Ras-dependent phosphorylation pathways regulate Myc protein stability. *Genes Dev.* **14**, 2501–2514 (2000).
- Nikolakaki, E., Coffey, P. J., Hemelsoet, R., Woodgett, J. R. & Defize, L. H. Glycogen synthase kinase 3 phosphorylates Jun family members *in vitro* and negatively regulates their transactivating potential in intact cells. *Oncogene* **8**, 833–840 (1993).
- Trowbridge, J. J., Xenocostas, A., Moon, R. T. & Bhatia, M. Glycogen synthase kinase-3 is an *in vivo* regulator of hematopoietic stem cell repopulation. *Nature Med.* **12**, 89–98 (2006).
- Sato, N., Meijer, L., Skaltsounis, L., Greengard, P. & Brivanlou, A. H. Maintenance of pluripotency in human and mouse embryonic stem cells through activation of Wnt signaling by a pharmacological GSK-3-specific inhibitor. *Nature Med.* **10**, 55–63 (2004).
- Ayton, P. M. & Cleary, M. L. Molecular mechanisms of leukemogenesis mediated by *MLL* fusion proteins. *Oncogene* **20**, 5695–5707 (2001).
- Hess, J. L. *MLL*: a histone methyltransferase disrupted in leukemia. *Trends Mol. Med.* **10**, 500–507 (2004).
- Lavau, C., Szilvassy, S. J., Slany, R. & Cleary, M. L. Immortalization and leukemic transformation of a myelomonocytic precursor by retrovirally transduced HRX-ENL. *EMBO J.* **16**, 4226–4237 (1997).
- Somervaille, T. C. & Cleary, M. L. Identification and characterization of leukemia stem cells in murine *MLL-AF9* acute myeloid leukemia. *Cancer Cell* **10**, 257–268 (2006).
- Watake, K. *et al.* Lithium therapy improves neurological function and hippocampal dendritic arborization in a spinocerebellar ataxia type 1 mouse model. *PLoS Med.* **4**, e182 (2007).
- Polakis, P. The oncogenic activation of beta-catenin. *Curr. Opin. Genet. Dev.* **9**, 15–21 (1999).
- Testa, J. R. & Tsichlis, P. N. AKT signaling in normal and malignant cells. *Oncogene* **24**, 7391–7393 (2005).
- Nickeleit, I., Zender, S., Kossatz, U. & Malek, N. P. p27^{kip1}: a target for tumor therapies? *Cell Div.* **2**, 13 (2007).
- Milne, T. A. *et al.* Menin and *MLL* cooperatively regulate expression of cyclin-dependent kinase inhibitors. *Proc. Natl Acad. Sci. USA* **102**, 749–754 (2005).
- Yokoyama, A. *et al.* The menin tumor suppressor protein is an essential oncogenic cofactor for *MLL*-associated leukemogenesis. *Cell* **123**, 207–218 (2005).
- Hughes, C. M. *et al.* Menin associates with a trithorax family histone methyltransferase complex and with the *Hoxc8* locus. *Mol. Cell* **13**, 587–597 (2004).
- Xia, Z. B. *et al.* The *MLL* fusion gene, *MLL-AF4*, regulates cyclin-dependent kinase inhibitor *CDKN1B* (p27^{kip1}) expression. *Proc. Natl Acad. Sci. USA* **102**, 14028–14033 (2005).
- Surjit, M. & Lal, S. K. Glycogen synthase kinase-3 phosphorylates and regulates the stability of p27^{kip1} protein. *Cell Cycle* **6**, 580–588 (2007).
- G.-Amlak, M. *et al.* Regulation of myeloma cell growth through Akt/Gsk3/forkhead signaling pathway. *Biochem. Biophys. Res. Commun.* **297**, 760–764 (2002).
- Dimartino, J. F. & Cleary, M. L. *MLL* rearrangements in hematological malignancies: lessons from clinical and biological studies. *Br. J. Haematol.* **106**, 614–626 (1999).
- Chen, C. S. *et al.* Molecular rearrangements on chromosome 11q23 predominate in infant acute lymphoblastic leukemia and are associated with specific biologic variables and poor outcome. *Blood* **81**, 2386–2393 (1993).
- Gould, T. D. & Manji, H. K. The Wnt signaling pathway in bipolar disorder. *Neuroscientist* **8**, 497–511 (2002).
- Yilmaz, O. H. *et al.* *Pten* dependence distinguishes haematopoietic stem cells from leukaemia-initiating cells. *Nature* **441**, 475–482 (2006).
- Smith, K. S., Rhee, J. W. & Cleary, M. L. Transformation of bone marrow B-cell progenitors by E2A-HIF requires coexpression of BCL-2. *Mol. Cell. Biol.* **22**, 7678–7687 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank R. Roth for providing AKT constructs, P. J. Roach for providing GSK3 constructs, D. G. Gilliland for providing a TEL-AML1 construct, M. Iwasaki for NUP98-HOXA9 cells, M. Ambros and C. Nicolas for technical assistance, and members of the Cleary laboratory for discussions. We acknowledge support from the Children's Health Initiative of the Packard Foundation, PHS grants CA55029 and CA116606, the Leukemia and Lymphoma Society, the Williams Lawrence Foundation and a Developmental Research Award from the Stanford Cancer Center.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to M.L.C. (mcleary@stanford.edu).

METHODS

Mice. C57BL/6 and SCID mice were obtained from the breeding facility of the Stanford University Veterinary Service Center. *Gsk3b*^{+/-} mice (provided by G. R. Crabtree with permission from J. R. Woodgett) were maintained on a CD1 genetic background. All experiments on mice were performed with the approval and in accordance with Stanford's Administrative Panel on Laboratory Animal Care.

Inhibitors. The GSK3 inhibitors SB216763 (Sigma), GSK3-IX and alsterpaullone (EMDbiosciences) were dissolved in dimethylsulphoxide and used at the indicated concentrations. All other inhibitors (EMDbiosciences) were dissolved in dimethylsulphoxide and used at the concentrations indicated in Supplementary Table 1.

Cell cultures. All human leukaemia cell lines (Supplementary Table 2) were maintained in R10 medium (RPMI1640 supplemented with 10% FBS, 1% L-glutamine and penicillin/streptomycin). Immortalized mouse myeloid cells were maintained in R20/20 medium (RPMI1640 with 20% FCS, 20% WEHI-conditioned medium, 1% L-glutamine and penicillin/streptomycin). Immortalized mouse B cells were cultured in OP9 medium (MEM α +GlutaMax 1 with 10% FBS, 1% L-glutamine, penicillin/streptomycin and 2 μ M β -mercaptoethanol) containing 1 ng ml⁻¹ IL-7 when necessary. All culture medium was obtained from Gibco.

DNA constructs and virus production. Retroviral constructs (MSCV vector) encoding *MLL-ENL*, *MLL-LAF4*, *MLL-AF6*, *MLL-GAS7*, *E2A-PBX1*, *NUP98-HOXA9* and *E2A-HLF* were reported previously^{21,28,37-40}. Retroviral constructs encoding *CA-AKT*, *ER-CA-AKT*⁴¹, and wild-type or S9A mutant GSK3B were constructed by cloning the respective complementary DNAs into MSCV using standard cloning techniques. Retrovirus production was performed as described previously⁴². Oligonucleotides for specific shRNA knockdown of *Gsk3a*, *Gsk3b* or *p27^{Kip1}* (sequences in Supplementary Table 3) were designed using PSICOLIGOMAKER 1.5 software (http://web.mit.edu/jacks-lab/protocols_table.html), and cloned into pSicoR-Puro or pSicoR-Hygromycin lentiviral vectors. Lentiviral stocks were produced as described previously⁴³.

Murine progenitor transformation assays. Myeloid progenitor transformation was performed as described previously²¹ with minor modifications. In brief, c-Kit⁺ cells were isolated from the bone marrow of 4–8-week-old C57BL/6 mice or E16 fetal livers (*Gsk3b*^{-/-} mice) using an auto-MACS and anti-c-Kit beads (Miltenyi Biotech). The c-Kit⁺ cells were spinoculated with retroviral supernatant in the presence of 5 μ g ml⁻¹ polybrene for 2 h at 1,350g and at 32 °C. After spinoculation and after overnight culture, cells were plated in methylcellulose medium (M3231; Stem Cell Technologies) containing 20 ng ml⁻¹ stem cell factor, 10 ng ml⁻¹ IL-6, granulocyte-macrophage colony-stimulating factor (GM-CSF), and IL-3 (R&D Systems) with appropriate antibiotic selection. After 5–7 days of culture, colonies were counted, pooled, and then 10⁴ cells were replated in the same medium but without antibiotic. At the end of the fourth round, cells were transferred to R20/20 medium to establish continuous cell lines.

B cell progenitors were transduced as described previously³⁷ with minor modifications. Transduced cells were co-cultured on neo-resistant-irradiated OP9 stromal cells. After continuous passage and adaptation to liquid culture, immortalized cell lines were used for injections of syngeneic recipient mice. *MLL-AF4* B cell precursor leukaemia cell lines were generated by explantation of splenocytes collected from leukaemic mice.

Transduction of immortalized mouse cells. Immortalized mouse cells (20,000) were transduced with retroviral or lentiviral constructs by spinoculation at

2,500g or 1,200g for 2 h at 32 °C. Transduced cells were then resuspended in 200 μ l of R20/20 or OP9 medium and transferred to 96-well plates. After overnight incubation at 37 °C, myeloid cells were plated in methylcellulose medium containing IL-3, IL-6, GM-CSF and stem cell factor. Transduced B cells were plated in methylcellulose medium containing IL-7.

Leukaemogenesis assays. Myeloid progenitors (wild type or *Gsk3b*^{-/-}) transformed by *MLL-ENL* were transduced with different lentiviruses, selected for drug resistance and then transplanted (10⁶ cells) by intravenous injection into sub-lethally irradiated (2 Gy) C.B-17 *scid/scid* mice (6–8-weeks-old). For lithium treatment, irradiated (1 Gy) C57BL/6 mice were transplanted with *MLL-AF4*-transformed B cell progenitors (50,000 cells) and maintained on 0.4% lithium-carbonate-containing chow with saline water (Harlan Teklad). Lithium treatment was initiated 3 days before transplantation and continued for 30 days, at which point treatment was withheld for 5 days to allow recovery from drug-induced diuresis, and was then resumed. Development of acute leukaemia was confirmed by blood smear, peripheral blood leukocyte counts, FACS analyses and/or histology.

Flow cytometry. Staining of cells for FACS analysis was performed as previously described⁴² using conjugated antibodies obtained from either BD Pharmingen or eBioscience. Cell cycle assays using propidium iodide staining, and apoptosis assays using annexin V staining, were performed as described²⁸. BrdU incorporation was determined using the BrdU flow kit (BD Pharmingen) according to the manufacturer's instructions.

Cell proliferation and MTT assays. Cultured cells (10,000–20,000) were plated in 96-well plates in R10, R20/20 or OP9 medium (100 μ l) containing different concentrations of the indicated kinase inhibitors (Supplementary Table 1) and incubated at 37 °C. The numbers of viable cells were determined by trypan-blue dye exclusion at the indicated times using a haemocytometer. For MTT assays, cells were cultured for 3–4 days and then quantified using a cell proliferation kit 1 under conditions recommended by the manufacturer (Roche).

Western blot. Cells were lysed in buffer A (20 mM Tris, pH 7.5, 150 mM NaCl, 1% Nonidet P-40, 1 mM EDTA) containing protease inhibitors (complete mini protease inhibitors; Roche). Proteins (40 μ g) were subjected to SDS-PAGE and analysed by western blot using primary antibodies specific for GSK3 (Upstate Biotechnology), β -catenin (Upstate Biotechnology), phosph-GSK3 (Cell Signalling), AKT (Cell Signaling), tubulin (Sigma), p27 or p21 (Santa Cruz Biotechnology).

38. So, C. W. *et al.* MLL-GAS7 transforms multipotent hematopoietic progenitors and induces mixed lineage leukemias in mice. *Cancer Cell* 3, 161–171 (2003).
39. Smith, K. S., Jacobs, Y., Chang, C. P. & Cleary, M. L. Chimeric oncoprotein E2a-Pbx1 induces apoptosis of hematopoietic cells by a p53-independent mechanism that is suppressed by Bcl-2. *Oncogene* 14, 2917–2926 (1997).
40. Kasper, L. H. *et al.* CREB binding protein interacts with nucleoporin-specific FG repeats that activate transcription and mediate NUP98-HOXA9 oncogenicity. *Mol. Cell. Biol.* 19, 764–776 (1999).
41. Kohn, A. D. *et al.* Construction and characterization of a conditionally active version of the serine/threonine kinase Akt. *J. Biol. Chem.* 273, 11937–11943 (1998).
42. So, C. W. & Cleary, M. L. MLL-AFX requires the transcriptional effector domains of AFX to transform myeloid progenitors and transdominantly interfere with forkhead protein function. *Mol. Cell. Biol.* 22, 6542–6552 (2002).
43. Ventura, A. *et al.* Cre-lox-regulated conditional RNA interference from transgenes. *Proc. Natl Acad. Sci. USA* 101, 10380–10385 (2004).

ARTICLES

TMEM16A confers receptor-activated calcium-dependent chloride conductance

Young Duk Yang¹, Hawon Cho¹, Jae Yeon Koo¹, Min Ho Tak¹, Yeongyo Cho¹, Won-Sik Shim¹, Seung Pyo Park¹, Jesun Lee¹, Byeongjun Lee¹, Byung-Moon Kim¹, Ramin Raouf², Young Ki Shin¹ & Uhtaek Oh¹

Calcium (Ca^{2+})-activated chloride channels are fundamental mediators in numerous physiological processes including transepithelial secretion, cardiac and neuronal excitation, sensory transduction, smooth muscle contraction and fertilization. Despite their physiological importance, their molecular identity has remained largely unknown. Here we show that transmembrane protein 16A (TMEM16A, which we also call anoctamin 1 (ANO1)) is a bona fide Ca^{2+} -activated chloride channel that is activated by intracellular Ca^{2+} and Ca^{2+} -mobilizing stimuli. With eight putative transmembrane domains and no apparent similarity to previously characterized channels, ANO1 defines a new family of ionic channels. The biophysical properties as well as the pharmacological profile of ANO1 are in full agreement with native Ca^{2+} -activated chloride currents. ANO1 is expressed in various secretory epithelia, the retina and sensory neurons. Furthermore, knockdown of mouse *Ano1* markedly reduced native Ca^{2+} -activated chloride currents as well as saliva production in mice. We conclude that ANO1 is a candidate Ca^{2+} -activated chloride channel that mediates receptor-activated chloride currents in diverse physiological processes.

Chloride (Cl^-) channels have fundamental roles in many physiological functions. Members of one group of these channels are activated by intracellular Ca^{2+} , and, accordingly, are collectively referred to as Ca^{2+} -activated chloride channels (CaCCs). CaCCs control the apical outflux of Cl^- , which is essential for the vectorial transport of electrolytes and water by means of secretory epithelia in kidneys, airways, intestine, pancreas and salivary glands^{1–3}. In addition, CaCCs are also known to contribute to cardiovascular functions by controlling vascular endothelial as well as smooth muscle tone and by regulating cardiac myocyte excitability^{4,5}. Moreover, by changing membrane potentials, CaCCs also regulate neuronal cell excitability, which is responsible for controlling phototransduction, olfaction, gustation and somesthetic sensations².

The CaCC-mediated cellular responses are initiated by the stimulation of G-protein-coupled receptors (GPCRs). Many GPCR ligands such as ATP, acetylcholine, endothelin-1, angiotensin II, lysophosphatidic acid and histamine can evoke Ca^{2+} -dependent chloride currents⁴. Interestingly, despite their wide expression pattern, endogenous CaCCs exhibit very similar characteristic biophysical and pharmacological properties: Ca^{2+} -activated Cl^- currents are sensitive to Cl^- channel blockers, show small channel conductances and have a voltage-dependent Ca^{2+} sensitivity. Furthermore, their ion permeability follows the selectivity sequence $\text{I}^- > \text{Br}^- > \text{Cl}^- > \text{F}^-$ (refs 1, 6, 7).

Because of their pathophysiological significance, many efforts have been made to clone CaCCs. CLCA1, isolated from bovine trachea, bestrophins related to Best disease and human CLC3 have been identified as candidate CaCCs^{8–10}. Although these candidate genes exhibit some of the properties of the native CaCCs, incontrovertible evidence of their involvement in native Ca^{2+} -activated chloride currents is lacking or disputed^{1,3,11}. Here we characterize a new family of chloride channels, exemplified by ANO1, and provide evidence for their direct involvement in mediating endogenous Ca^{2+} -activated chloride current.

Characterization of mouse ANO1

In our attempts to unravel the molecular identity of endogenous CaCCs, we searched public domain databases for putative channel- or

transporter-like genes with more than two transmembrane domains and multiple isoforms. Amongst putative transmembrane proteins with unknown functions, TMEM16A attracted our attention because it has multiple putative membrane-spanning domains and ten homologues in man (Fig. 1a). Because the expressed sequence tags of TMEM16A are frequently found in retina, full-length *Tmem16a* complementary DNA was amplified using polymerase chain reaction with reverse transcription (RT-PCR) from total RNA isolated from mouse eye. Mouse *Tmem16a* cDNA contains an open reading frame of 2,880 nucleotides that encodes a protein of 960 amino acids (Supplementary Fig. 1). Mouse TMEM16A has high homology (91% amino-acid sequence identity) with its human orthologue, and its predicted rat orthologue has an additional 123 amino acids at its amino terminus; the remainder of the rat sequence has high homology with mouse TMEM16A (99% identity). Interestingly, the TMEM16 family does not share significant sequence homology with other families of ion channels.

Hydropathy analysis suggests that TMEM16A has eight transmembrane-spanning domains (Fig. 1b). Consensus analysis predicted that ANO1 has multiple protein kinase A, protein kinase C, protein kinase G and casein kinase phosphorylation sites at intracellular protein segments and multiple glycosylation sites in extracellular segments. Even though TMEM16A is activated by intracellular Ca^{2+} (see below), no specific consensus sites for Ca^{2+} binding, such as the EF hand of voltage-gated Ca^{2+} channels or the IQ motif of Ca^{2+} -calmodulin, were found. Because TMEM16A has eight transmembrane domains and an anionic channel property (see below), we tentatively called it anoctamin 1 (ANO1).

ANO1 mediates a GPCR-activated chloride current

Under physiological conditions CaCCs are activated by GPCR stimulation. Endothelin-1 is a powerful vasoconstrictor that depolarizes vascular smooth muscle by activating CaCCs⁴. We expressed mouse ANO1 and endothelin receptor subtype A ($\text{ET}_\text{A}\text{R}$, also known as EDNRA) in HEK 293T cells (ANO1/ $\text{ET}_\text{A}\text{R}$ -HEK) along with green

¹Sensory Research Center, CRI, College of Pharmacy, Seoul National University, Seoul 151-742, Korea. ²Molecular Nociception Group, Neuroscience, Physiology and Pharmacology, University College London, London WC1E 6BT, UK.

fluorescent protein (GFP) for visual identification of transfected cells, and tested the current response to endothelin-1 stimulation in whole-cell configuration. Pipette and bath solutions contained 140 mM *N*-methyl *D*-glucamine (NMDG)-Cl, which permitted Cl^- to be used as the main charge carrier. The pipette solution contained no added Ca^{2+} or EGTA. Under these conditions, application of 50 nM endothelin-1 to ANO1/ ET_A R-HEK cells evoked robust inward currents when the membrane potential was held at -60 mV (Fig. 1c). In contrast, endothelin-1-evoked currents were minimal in HEK 293T cells transfected with ET_A R only or control (in which only GFP was transfected) cells. Similarly, the co-expression in HEK 293T cells of ANO1 with rat angiotensin II receptor subtype 1 (AT1R, also known as AGTR1A), muscarinic receptor subtype 1 (M1R, also known as CHRM1), histamine receptor subtype 1 (H1R, also known as HRH1) or purinergic receptor subtype 2 (P2Y2R) (all of which are known to use native CaCCs for their cellular functions)^{12–14} resulted in robust current responses to angiotensin II, carbachol, histamine and ATP, respectively (Fig. 1c). No significant currents were observed in control transfectants. Moreover, in ANO1-transfected cells, the application of endothelin-1, ATP or carbachol evoked small but significantly greater currents than in control cells, presumably owing to the presence of endogenous receptors for these ligands in control HEK 293T cells (Fig. 1c). These results demonstrate that ANO1 can mediate

receptor-activated chloride currents in HEK 293T cells in a similar manner to endogenously activated CaCCs.

ANO1 recapitulates the properties of endogenous CaCCs

We next examined the current–voltage (*I/V*) relationship of endothelin-1-induced ANO1 currents (Fig. 2a). Voltage ramps from -80 to $+80$ mV were delivered to ANO1/ ET_A R-HEK cells. We found that during the early phase of the response, where the developing current was small, an outward-rectifying *I/V* relationship was observed. In contrast, a linear *I/V* relationship was observed during the peak current response. This is consistent with the linear or outwardly rectifying *I/V* relationships of endogenous CaCCs when intracellular Ca^{2+} concentrations are high or low^{6,15,16}. To determine the ionic selectivity of ANO1, voltage ramps from -80 to $+80$ mV were delivered to ANO1/ ET_A R-HEK cells. Endothelin-1 (50 nM) was applied to whole cells to evoke ANO1 currents; the pipette solution contained 140 mM NMDG-Cl (0 mM Ca^{2+} , 0 mM EGTA). When the bath solution contained equimolar NMDG-Cl, a linear *I/V* relationship that reversed near 0 mV (-4.5 mV) was observed (Fig. 2b). Furthermore, when the bath solution was changed to 140 mM sodium gluconate, no outward current was observed. Because NMDG⁺ and gluconate are incapable of passing through channels, these results demonstrate that ANO1 is an anion channel.

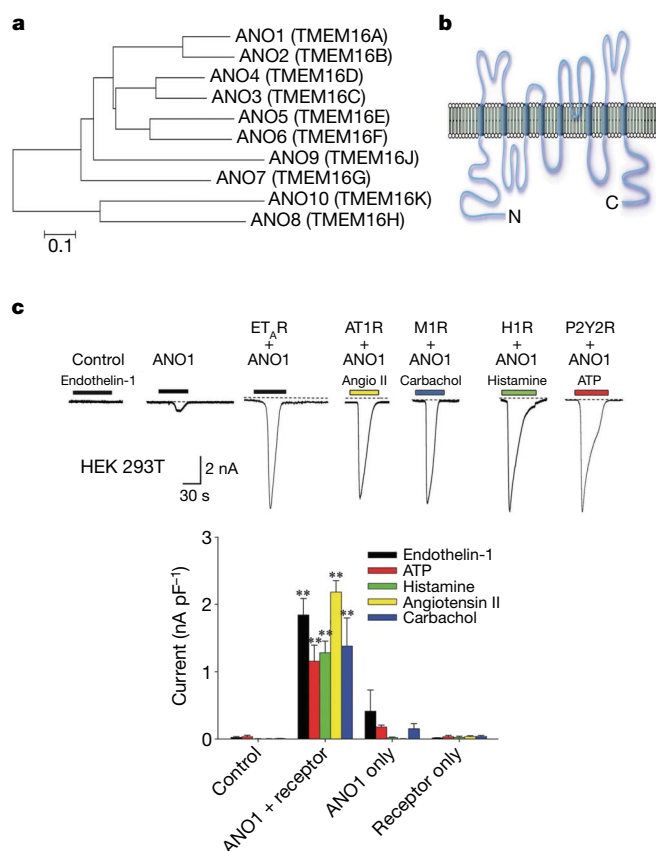


Figure 1 | ANO1-transfected cells respond with robust inward currents to GPCR stimulation. **a**, Phylogenetic tree depicting the anoctamin family in human. Scale bar, 0.1 nucleotide substitutions per site. **b**, Predicted topology of mouse ANO1. **c**, Top panel: large inward currents were activated by endothelin-1 (50 nM), angiotensin II (1 μM , Angio II), carbachol (1 μM), histamine (100 μM) or ATP (100 μM) when mouse ANO1 was co-transfected into HEK 293T cells with ET_A R, AT1R, M1R, H1R or P2Y2R. $E_\text{hold} = -60$ mV. Dashed lines represent zero current. Lower panel: summary of receptor-stimulation-activated currents (experiment numbers = 5–18). ** $P < 0.001$ compared to control (analysis of variance (ANOVA), Duncan's post-hoc test). Error bars, mean \pm s.e.m.

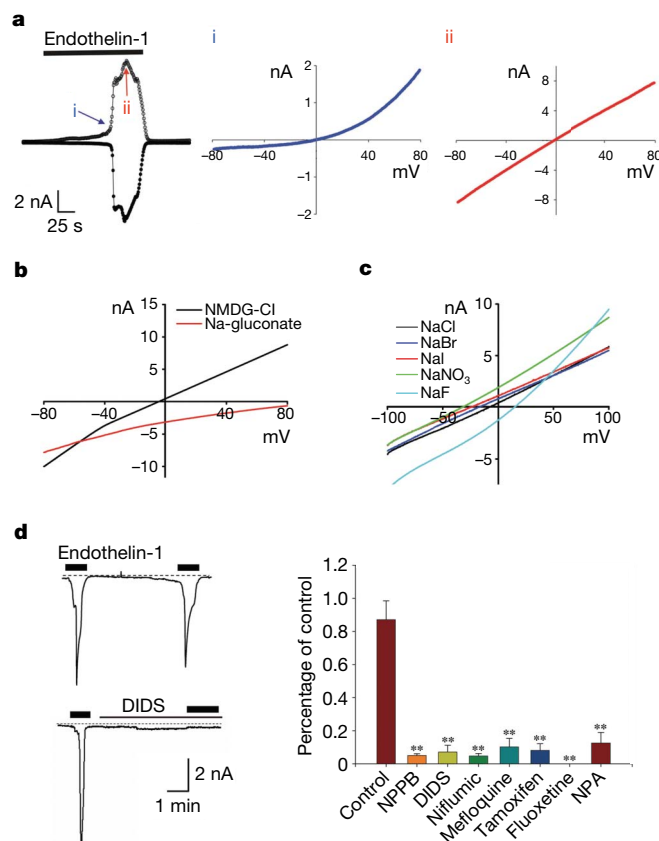


Figure 2 | ANO1 is a Cl^- channel. **a**, Typical current–voltage relationships of endothelin-1-evoked currents at different time points in ANO1/ ET_A R-HEK cells. Voltage ramps: -80 to $+80$ mV, 320 ms duration, 0.83 Hz. **b**, Anion selectivity of endothelin-1-evoked currents in ANO1/ ET_A R-HEK cells. Pipette solution contained 140 mM NMDG-Cl. Bath solutions contained 140 mM NMDG-Cl or Na-gluconate. **c**, Anion permeability. Bath solution contained 140 mM NaCl, NaBr, NaI, NaNO_3 or NaF. **d**, Inhibition of endothelin-1-evoked currents in ANO1/ ET_A R-HEK cells by Cl^- channel blockers as well as classical inhibitors ($n = 5$ for each) of native CaCCs. 10 μM of chemicals (except mefloquine, 5 μM) were applied 5 min before the second application of endothelin-1. Control ($n = 12$). NPA, *n*-phenylanthranilic acid. ** $P < 0.001$ compared to control (ANOVA, Duncan's post-hoc test). Error bars, mean \pm s.e.m.

The permeability sequence of monovalent anions is characteristic of CaCCs^{2,4,17}, and, thus, we investigated the relative permeability of ANO1 to monovalent anions. For this experiment the pipette solution contained 140 mM NMDG-Cl (0 mM Ca^{2+} , 0 mM EGTA). Permeability ratios (P_X/P_{Cl}) were estimated using the Goldman-Hodgkin and Katz equation from reversal potential changes induced by replacing extracellular Cl^- with NO_3^- , Br^- , I^- or F^- ¹⁸. When NaCl in the bath was substituted with equimolar NaF, NaBr, NaI or NaNO_3 , reversal potentials shifted from -9.0 ± 0.6 mV (mean \pm s.e.m.; $n = 5$) to $+12.5 \pm 2.9$ mV ($n = 8$), -23.1 ± 2.2 mV ($n = 5$), -24.7 ± 2.5 mV ($n = 5$) and -29.1 ± 1.1 mV ($n = 7$), respectively (Fig. 2c). Thus, the relative permeabilities of these monovalent anions were NO_3^- (2.20) $>$ I^- (1.85) $>$ Br^- (1.74) $>$ Cl^- (1.0) $>$ F^- (0.43), which is entirely consistent with those of endogenous CaCCs (Supplementary Table 1)^{2,4,17}.

Agonist-induced currents in ANO1-expressing cells were blocked by Cl^- channel blockers, 4,4'-diisothiocyanatostilbene-2,2'-disulphonic acid (DIDS), niflumic acid or 5-nitro-2-(3-phenylpropylamino)-benzoic acid (NPPB)^{2,5,17}. In ANO1/ET_AR-HEK cells, 50 nM endothelin-1 was applied twice with a 5 min interval, and second responses were found to reach $87.1 \pm 11.4\%$ ($n = 12$) of first responses, eliciting weak tachyphylaxis. Pretreatment or co-application of DIDS, niflumic acid and NPPB markedly inhibited this response of ANO1/ET_AR-HEK cells to the second application of endothelin-1 (Fig. 2d). In addition, modifiers of various enzymes or transporters such as fluoxetine (a serotonin reuptake inhibitor), tamoxifen (an oestrogen receptor modulator), *n*-phenyl-anthranilic acid (a cyclooxygenase inhibitor) and mefloquine (an antimalarial drug) are known to block native CaCCs^{2,3,5}. If ANO1 is a candidate CaCC, mouse ANO1 should be inhibited by these CaCC blockers. Indeed, these chemicals markedly inhibited endothelin-1-induced currents in ANO1/ET_AR-HEK cells (Fig. 2d). This pharmacological profile further confirms that ANO1 is a CaCC candidate.

Mutation in the putative pore region of ANO1 changes ion permeability

We looked to determine whether ANO1 subunits directly assembled into an ion channel that has a pore and a selectivity filter. To determine whether ANO1 contains a pore region, we selected a highly conserved region by analogy with other homologues that lies between TM5 and TM6 consisting of 103 amino acids for mutation analysis (Supplementary Fig. 3). This region contains three positively charged residues (R621, K645 and K668) that are highly conserved among isoforms. We reasoned that mutating these positively charged residues would affect the ionic properties of the channel if the residues are part of the pore region. When the positively charged residues were replaced with a negatively charged residue (glutamate), a marked reduction in anion permeabilities and a notable increase in cation permeability were observed (Supplementary Information). This mutational study suggests that the region spanning TM5 and TM6 is a pore-forming region.

ANO1 is activated by intracellular Ca^{2+}

To prove that ANO1 is activated by Ca^{2+} , Ca^{2+} was applied to inside-out membrane patches isolated from ANO1/ET_AR-HEK cells. Ca^{2+} , applied to the bath (intracellular side) in various concentrations, evoked macroscopic single-channel currents in a dose-dependent manner ($E_{\text{hold}} = -60$ mV; Fig. 3a). A half-maximal concentration (EC_{50}) for ANO1 activation was $2.6 \mu\text{M}$ at -60 mV. The threshold concentration of Ca^{2+} reached about $0.3 \mu\text{M}$ at -60 mV holding potential (Fig. 3b). Remarkably, ANO1 activation by Ca^{2+} was voltage-dependent, which is also characteristic of endogenous CaCCs^{7,15,16}. At a depolarized membrane potential ($+60$ mV), ANO1 was more sensitive to Ca^{2+} and the concentration-response curve shifted to the left, resulting in an EC_{50} of $0.4 \mu\text{M}$ (Fig. 3b). Hill's coefficients of the concentration-response curves were 2.0 and 2.4 at -60 mV and $+60$ mV, respectively. Surprisingly, ANO1 current was activated less

by $[\text{Ca}^{2+}] > 10 \mu\text{M}$, which is another indication of CaCC candidacy because native CaCC current is also inhibited by high Ca^{2+} (ref. 19). When $30 \mu\text{M}$ Ca^{2+} was applied to inside-out patches from ANO1/ET_AR-HEK cells ($E_{\text{hold}} = -60$ mV) it reached to $53.3 \pm 17.1\%$ ($n = 5$) of the current amplitude induced by $10 \mu\text{M}$ Ca^{2+} .

The amplitudes of single-channel ANO1 currents were small, like those of endogenous CaCCs^{6,20}. At -60 mV, the average single-channel current amplitude was 0.42 ± 0.01 pA ($n = 6$; Fig. 3c). Moreover, we measured the single-channel current amplitudes of ANO1 activated by Ca^{2+} at various membrane potentials, and found a linear I/V relationship. In addition, the slope conductance of ANO1 was 8.3 pS, which is comparable to those of native CaCCs in various cell types^{6,20}.

Various GPCRs are known to activate CaCCs by means of the action of G α subunits of the G $_q$ class of G protein, which stimulates phospholipase C (PLC) to yield inositol-1,4,5-trisphosphate ($\text{Ins}(1,4,5)\text{P}_3$) and subsequent Ca^{2+} release from internal stores¹². Thus, it is likely that ANO1 activation by GPCRs is mediated by the PLC/ $\text{Ins}(1,4,5)\text{P}_3$ signalling pathway. Indeed, we observed that ANO1 is activated by intracellular signals that increase intracellular Ca^{2+} concentrations, most notably by means of the PLC/ $\text{Ins}(1,4,5)\text{P}_3$ pathway (Supplementary Information). ANO1 was also activated by

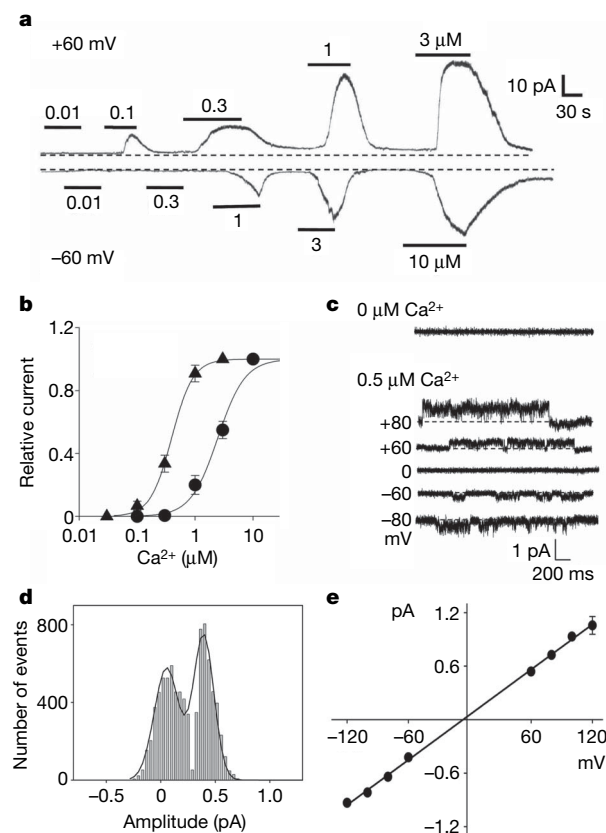


Figure 3 | ANO1 is activated by intracellular Ca^{2+} in a voltage-dependent manner. **a**, Macroscopic current responses to repeated applications of Ca^{2+} in various concentrations to the bath of inside-out patches isolated from ANO1/ET_AR-HEK cells at holding potentials of -60 or $+60$ mV. Pipette and bath solutions contained 140 mM NMDG-Cl. **b**, Dose-response relationship of ANO1 activation by Ca^{2+} . Current responses were normalized versus those observed at $10 \mu\text{M}$ (circles, -60 mV, $n = 8$) or $3 \mu\text{M}$ Ca^{2+} (triangles, $+60$ mV, $n = 8$). Data points were fitted to the Hill's equation. Error bars, mean \pm s.e.m. **c**, Single-channel currents activated by Ca^{2+} ($0.5 \mu\text{M}$) in an inside-out membrane patch. Signals were filtered at 2 kHz. **d**, An amplitude histogram of single-channel currents activated by Ca^{2+} at $+60$ mV. **e**, Current-voltage relationship of single-channel Cl^- currents activated by Ca^{2+} ($n = 3-7$). Error bars (mean \pm s.e.m.) are too small to be buried in symbols.

dialysing Ca^{2+} to whole cells as observed with endogenous CaCCs^{6,21} (Supplementary Information).

ANO1 expression pattern

In immunohistochemical analysis for visualization of tissue distribution, ANO1 was highly expressed in tissues in which endogenous CaCCs have been described. As shown in Fig. 4a, dense ANO1 immunoreactivity was observed in epithelial cells of pulmonary bronchioles, whereas expression in lung alveolar cells was less

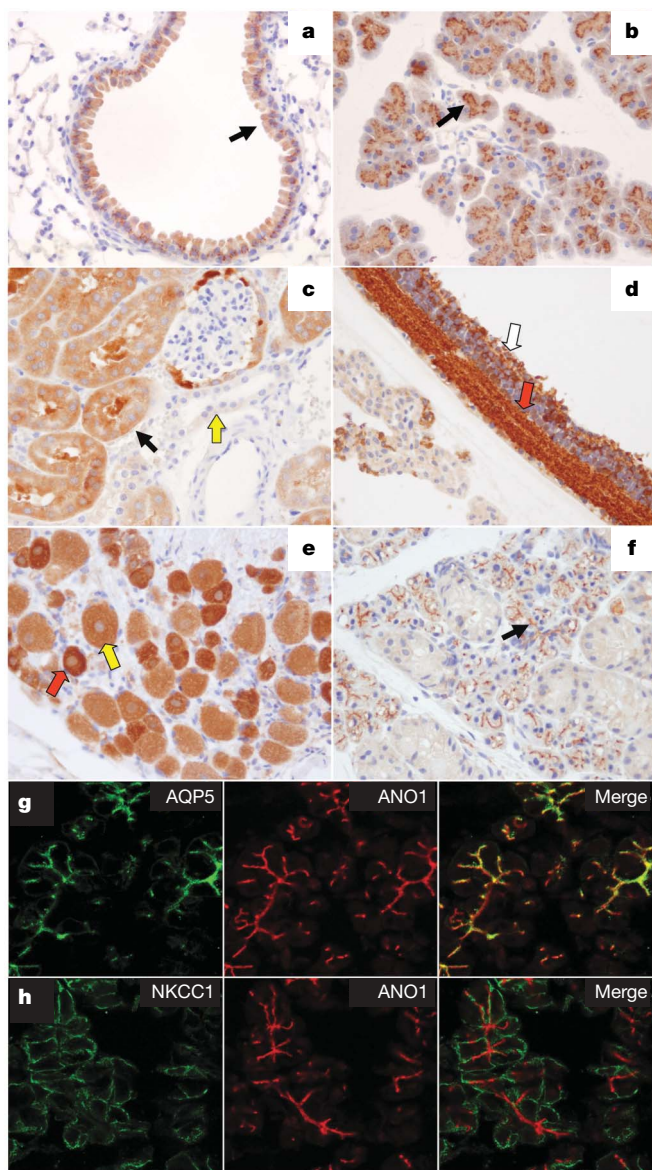


Figure 4 | ANO1 is expressed in transport epithelia and other tissues. Residues 451 to 466 of mouse ANO1 were chosen for generation of an ANO1 polyclonal antibody for immunohistochemical analysis. **a**, Lung, dense ANO1 immunoreactivity in epithelial cells in bronchioles (black arrow); **b**, pancreas acinar cells (black arrow); **c**, kidney, epithelia of proximal tubules (black arrow) and distal renal tubules (yellow arrow); **d**, retina, dense immunoreactivity in the outer nuclear (white arrow) and inner nuclear and ganglion cell layers (red arrow); **e**, dorsal-root ganglion, small (red arrow) and large (yellow arrow) sensory neurons; **f**, submandibular gland, ANO1 immunoreactivity in acinar cells (black arrow). **g, h**, To verify the expression of ANO1 in the apical (luminal) membrane of acinar cells, submandibular gland was double-stained with antibodies of mouse ANO1 and aquaporin 5 (AQP5, **g**) or NKCC1 (**h**), marker antibodies for the apical membrane or the basolateral membrane of acinar cells in submandibular gland. Magnification, $\times 400$.

prominent. Pancreatic acinar cells were also heavily stained by ANO1 antibody (Fig. 4b). In the kidney, dense immunoreactivity was observed mainly in the epithelia of proximal (black arrow) and weakly in distal (yellow arrow) renal tubules (Fig. 4c). Remarkably, as has been proven by electrophysiological experiments²², dense ANO1 immunoreactivity was found in all retinal cell layers (Fig. 4d), that is, the outer nuclear layer, which contains the bodies of photoreceptors (white arrow), and the inner nuclear and ganglion cell layers (red arrow). Most sensory neurons but not satellite cells in dorsal root ganglia stained positively for ANO1 (Fig. 4e), in agreement with the modulatory effects of CaCCs in sensory transmission²³. Small sensory neurons (red arrow) tended to stain more densely than large sensory neurons (yellow arrow), suggesting that ANO1 has a modulatory role in nociception, because small neurons in dorsal-root ganglia are intimately involved in nociception²⁴. In submandibular glands, dense immunoreactivity was observed at the apices of acinar cells colocalized with aquaporin 5, a marker for the apical membrane of acinar cells^{25,26} (Fig. 4g), but much less with NKCC1 (also known as SLC12A2), a marker for the basolateral membrane of acinar cells^{25,26} (Fig. 4h). In contrast, we did not find ANO1 immunoreactivity in the brain except for a weak immunoreactivity in neural tracts (data not shown). Furthermore, dense immunoreactivity was observed in Leydig cells and moderate immunoreactivity was observed in growing spermatocytes in testes (data not shown). Thus, the expression pattern of ANO1 in these secretory epithelial cells, retinal and sensory neurons coincides with the reported locations of CaCCs^{1,2}.

ANO1 mediates the physiological responses of CaCCs

Our biophysical and pharmacological evidence suggests that ANO1 confers Ca^{2+} -activated Cl^- conductance. To explore the physiological functions of ANO1, we investigated its relationship with the various functions of endogenous CaCCs. In particular, we investigated the role of ANO1 in salivary secretion because much evidence indicates that CaCCs are present in the acinar cells of salivary glands²¹. To address this issue, *Ano1* small interfering RNA (siRNA) was injected into mice intravenously (retro-orbital injection) to knockdown *Ano1*; scrambled siRNA was also administered as a control. To facilitate siRNA tissue uptake, it was mixed with a liposome-like material, polyethylenimine, before injection (see Methods). Four days after this siRNA administration, pilocarpine-induced saliva was collected to measure the rate of saliva production over 20 min, because the downregulation of *Ano1* was expected to reduce saliva production. Pilocarpine, a muscarinic receptor agonist, administration (1 mg kg^{-1} , intraperitoneally (i.p.)) greatly increased salivary production in control mice, which peaked at 15 min after injection. The salivary production was not affected by scrambled siRNA, whereas *Ano1* siRNA significantly reduced peak salivary flow induced by pilocarpine (Fig. 5b). Consistent with reduction in salivation, acetylcholine-induced currents in primary cultures of submandibular glands isolated from siRNA-injected mice were significantly smaller than those of control mice and scrambled siRNA-treated mice ($P < 0.05$, ANOVA, Duncan's post-hoc test; Fig. 5c). As expected, ANO1 expression in submandibular glands was greatly reduced in siRNA-treated mice compared to those of control or scrambled siRNA-injected animals (Fig. 5a). The osmolarities of saliva and concentrations of Na^+ , K^+ and Cl^- in saliva samples of siRNA, scrambled siRNA-treated and control mice were similar (Supplementary Table 2). These results suggest that ANO1 is required for agonist-induced salivary secretion.

Discussion

Several candidate CaCC genes have been cloned. CLCA1 was initially isolated from bovine trachea¹⁰. Some CLCA phenotypes are similar to those of native CaCCs^{10,27}. However, because of the different biophysical properties of bovine CLCA2, CLCA does not well match the requirements of a CaCC^{1,3,11}. Moreover, bovine CLCA2 is now

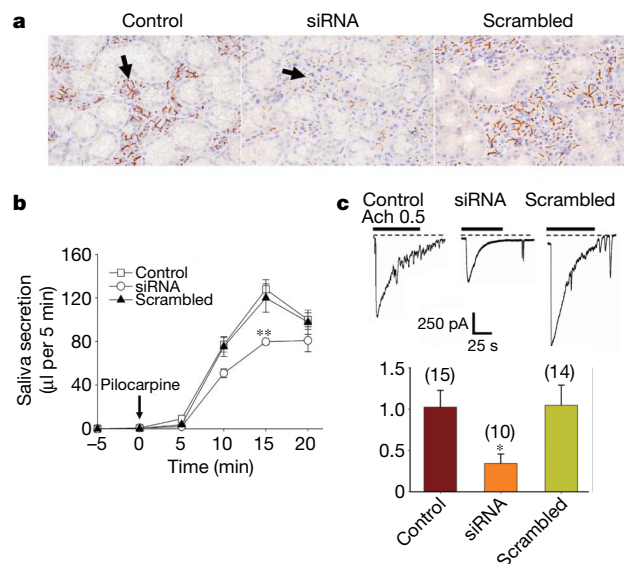


Figure 5 | Mouse *Ano1* siRNA reduces pilocarpine-induced salivary output. **a**, Expression of ANO1 (arrow) in the submandibular glands of control, *Ano1* siRNA-treated and scrambled siRNA-treated animals. **b**, Effects of *Ano1* siRNA treatment on saliva production. Saliva output was measured in every 5 min in control ($n = 10$), siRNA-treated ($n = 8$) and scrambled siRNA-treated ($n = 10$) mice. Pilocarpine (1 mg kg^{-1}) was injected intraperitoneally (arrow) to induce sufficient saliva production. $**P < 0.01$ compared to control (ANOVA, Duncan's post-hoc test). Error bars, mean \pm s.e.m. **c**, Upper panel: representative traces of acetylcholine (ACh, $0.5 \mu\text{M}$)-induced currents in cells of submandibular glands isolated from mice injected with *Ano1* siRNA or scrambled RNA four days before culture. Lower panel: summary of current responses of acinar cells in submandibular glands. $*P < 0.05$ compared to control (ANOVA, Duncan's post-hoc test). Error bars, mean \pm s.e.m.

considered to be an endothelial adhesion molecule that mediates tumour cell adhesion²⁸. Human CLCA3 seems to be a secreted protein and to have no transmembrane domain after truncation²⁹. Notably, mouse CLCA1 produces Ca^{2+} -activated Cl^- currents that resemble those of endogenous CaCCs when co-expressed with KCNB1 (a regulatory subunit of Ca^{2+} -activated K^+ channel)³⁰. The bestrophins are proteins related to Best disease, a vitelliform macular dystrophy⁸. These proteins have biophysical properties that are similar to those of endogenous CaCCs, that is, direct activation by intracellular Ca^{2+} at physiological concentrations and the anion permeability sequence⁸. In addition, mutation in amino acids alters anion permeabilities³¹. Knockdown of bestrophin 1 reduces Ca^{2+} -activated Cl^- conductances³². CLC3 (a homologue of the voltage-gated chloride channel CLC-0) has also been suggested to be a CaCC⁹. However, the general biophysical properties of CLC3 differ substantially from those of native CaCCs. *Clc3*^{-/-} mice show Ca^{2+} -activated Cl^- conductance³³. These channels may therefore contribute to aspects of calcium-activated chloride flux, but other channels are also likely to be involved in this phenomenon.

Cystic fibrosis is a life-threatening disease, in which Cl^- secretion by airway and other secretory epithelia is reduced³⁴. Cystic fibrosis transmembrane regulator (CFTR), a Cl^- channel activated by intracellular cAMP, is responsible for the dysfunction in Cl^- secretion in cystic fibrosis patients^{34,35}. However, compelling evidence suggests that CaCCs are also involved in Cl^- secretion in these tissues. In mice, CaCCs appear to compensate for CFTR loss of function; this is shown by no reduction in fluid secretion or apparent cystic fibrosis symptoms in CFTR-deficient mice³⁶. Even though such functional compensation is not evident in human cystic fibrosis, evidence indicates that the activation of latent CaCCs could compensate for CFTR loss of function in cystic fibrosis patients^{37,38}. The treatment of

human airway epithelial cells with UTP, a purinergic receptor agonist, was found to cause Cl^- conductance in cystic fibrosis epithelia³⁹. These results suggest that CaCC activation could compensate for CFTR loss of function in cystic fibrosis, and that ANO1 may present an alternative target in cystic fibrosis.

Recently, evidence has been presented indicating that ANO1 proteins participate in tumorigenesis. ANO1 is located in the 11q13 human chromosome, where many tumour-related genes such as cyclin D1 are clustered⁴⁰. ANO1 is markedly upregulated in squamous cell carcinoma of the head and neck⁴¹ and in gastrointestinal stromal tumours⁴². The reason for this propensity of ANO1 in tumours is unknown, but it is conceivable that a secretory environment might be necessary for tumour cell proliferation. The abundance of ANO1 in tumours may provide clues relating to the treatment or prediction of cancer in these tissues.

ANO1 is thus the prototypical member of a structurally unique new family of ion channels. Its properties and distribution are consistent with its identity as the long sought CaCC. Because such channels are implicated in a wide range of physiological functions, gene deletion studies should prove informative. In addition, these channels may represent a new class of drug targets for a variety of diseases, ranging from cystic fibrosis to cancer. Further characterization of other members of the family should help define the significance of these new ion channels in health and disease.

METHODS SUMMARY

Cloning of mouse ANO1. First-strand cDNAs were reverse transcribed from total RNA isolated from mouse eye. The full-length coding sequence of *Ano1* (*Tmem16a*; NM_178642.4 gi: 110835695) was amplified by PCR, and subcloned into pSDF for oocyte expression or into pEGFP-N1 for the expression in mammalian cells.

Site-directed mutagenesis. All mutants of ANO1 were generated by site-directed mutagenesis, substituting the central 1–2 nucleotides of the desired mutagenic site with two complementary mutagenic primers using the Muta-Direct site-directed mutagenesis kit (iNtRON Biotech).

siRNA. The *Ano1* siRNAs and scrambled siRNA were synthesized and injected to the retroorbital sinuses in the orbits of 7-week-old male ICR mice with polyethyleneimine for efficient target delivery to tissues *in vivo*.

Salivary secretion test. Four days after siRNA injections, saliva was collected 5 min before and 5, 10, 15 and 20 min after administration of pilocarpine-HCl (1 mg kg^{-1} , i.p.). A piece of pre-weighed cotton ball ($\sim 10 \text{ mg}$) was inserted below each tongue to measure the saliva volumes. The saliva-soaked cotton balls were removed, weighed and then spun to extract saliva for measurements of osmolality and electrolyte content.

Current recordings. Whole cells were formed with a $\sim 3 \text{ M}\Omega$ Sylgard-coated glass pipette. The junction potential was adjusted to 0 mV. Whole-cell currents were amplified using an Axopatch 200B amplifier (Molecular Device) and stored in a personal computer after digitization by a Digidata 1440 (Molecular Device).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 16 May; accepted 5 August 2008.

Published online 24 August 2008.

- Hartzell, C., Putzier, I. & Arreola, J. Calcium-activated chloride channels. *Annu. Rev. Physiol.* **67**, 719–758 (2005).
- Frings, S., Reuter, D. & Kleene, S. J. Neuronal Ca^{2+} -activated Cl^- channels—homing in on an elusive channel species. *Prog. Neurobiol.* **60**, 247–289 (2000).
- Eggermont, J. Calcium-activated chloride channels: (un)known, (un)loved? *Proc. Am. Thorac. Soc.* **1**, 22–27 (2004).
- Large, W. A. & Wang, Q. Characteristics and physiological role of the Ca^{2+} -activated Cl^- conductance in smooth muscle. *Am. J. Physiol.* **271**, C435–C454 (1996).
- Nilius, B. & Droogmans, G. in *Calcium-Activated Chloride Channels* (ed. Fuller, C. M.) 327–344 (Academic, 2002).
- Nilius, B. et al. Calcium-activated chloride channels in bovine pulmonary artery endothelial cells. *J. Physiol. (Lond.)* **498**, 381–396 (1997).
- Nilius, B. et al. Kinetic and pharmacological properties of the calcium-activated chloride-current in macrovascular endothelial cells. *Cell Calcium* **22**, 53–63 (1997).
- Sun, H., Tsunenari, T., Yau, K. W. & Nathans, J. The vitelliform macular dystrophy protein defines a new family of chloride channels. *Proc. Natl Acad. Sci. USA* **99**, 4008–4013 (2002).

9. Huang, P. *et al.* Regulation of human CLC-3 channels by multifunctional Ca^{2+} /calmodulin-dependent protein kinase. *J. Biol. Chem.* **276**, 20093–20100 (2001).
10. Cunningham, S. A. *et al.* Cloning of an epithelial chloride channel from bovine trachea. *J. Biol. Chem.* **270**, 31016–31026 (1995).
11. Nilius, B. & Droogmans, G. Amazing chloride channels: an overview. *Acta Physiol. Scand.* **177**, 119–147 (2003).
12. Zholos, A. *et al.* Ca^{2+} - and volume-sensitive chloride currents are differentially regulated by agonists and store-operated Ca^{2+} entry. *J. Gen. Physiol.* **125**, 197–211 (2005).
13. Lee, M. G., Zeng, W. & Muallem, S. Characterization and localization of P2 receptors in rat submandibular gland acinar and duct cells. *J. Biol. Chem.* **272**, 32951–32955 (1997).
14. Guibert, C., Marthan, R. & Savineau, J. P. Oscillatory Cl^- current induced by angiotensin II in rat pulmonary arterial myocytes: Ca^{2+} dependence and physiological implication. *Cell Calcium* **21**, 421–429 (1997).
15. Kuruma, A. & Hartzell, H. C. Bimodal control of a Ca^{2+} -activated Cl^- channel by different Ca^{2+} signals. *J. Gen. Physiol.* **115**, 59–80 (2000).
16. Evans, M. G. & Marty, A. Calcium-dependent chloride currents in isolated cells from rat lacrimal glands. *J. Physiol. (Lond.)* **378**, 437–460 (1986).
17. Hartzell, C. *et al.* Looking chloride channels straight in the eye: bestrophins, lipofuscinosis, and retinal degeneration. *Physiology (Bethesda)* **20**, 292–302 (2005).
18. Arreola, J. & Melvin, J. E. A novel chloride conductance activated by extracellular ATP in mouse parotid acinar cells. *J. Physiol. (Lond.)* **547**, 197–208 (2003).
19. Fuller, C. M., Ismailov, I. I., Keeton, D. A. & Benos, D. J. Phosphorylation and activation of a bovine tracheal anion channel by Ca^{2+} /calmodulin-dependent protein kinase II. *J. Biol. Chem.* **269**, 26642–26650 (1994).
20. Piper, A. S. & Large, W. A. Multiple conductance states of single Ca^{2+} -activated Cl^- channels in rabbit pulmonary artery smooth muscle cells. *J. Physiol. (Lond.)* **547**, 181–196 (2003).
21. Arreola, J., Melvin, J. E. & Begenisich, T. Activation of calcium-dependent chloride channels in rat parotid acinar cells. *J. Gen. Physiol.* **108**, 35–47 (1996).
22. Maricq, A. V. & Korenbrot, J. I. Calcium and calcium-dependent chloride currents generate action potentials in solitary cone photoreceptors. *Neuron* **1**, 503–515 (1988).
23. Kenyon, J. L. & Scott, R. H. in *Calcium-Activated Chloride Channels* (ed. Fuller, C. M.) 135–166 (Academic, 2002).
24. Willis, W. D. & Coggeshall, R. E. *Sensory Mechanisms of the Spinal Cord* (eds Willis, W. D. & Coggeshall, R. E.) (Plenum, 2004).
25. Liu, X. *et al.* Attenuation of store-operated Ca^{2+} current impairs salivary gland fluid secretion in $\text{TRPC1}^{-/-}$ mice. *Proc. Natl Acad. Sci. USA* **104**, 17542–17547 (2007).
26. He, X. *et al.* Polarized distribution of key membrane transport proteins in the rat submandibular gland. *Pflugers Arch.* **433**, 260–268 (1997).
27. Gruber, A. D. *et al.* Genomic cloning, molecular characterization, and functional analysis of human CLCA1, the first human member of the family of Ca^{2+} -activated Cl^- channel proteins. *Genomics* **54**, 200–214 (1998).
28. Elble, R. C. & Pauli, B. U. Tumor suppression by a proapoptotic calcium-activated chloride channel in mammary epithelium. *J. Biol. Chem.* **276**, 40510–40517 (2001).
29. Gruber, A. D., Schreier, K. D., Ji, H. L., Fuller, C. M. & Pauli, B. U. Molecular cloning and transmembrane structure of hCLCA2 from human lung, trachea, and mammary gland. *Am. J. Physiol.* **276**, C1261–C1270 (1999).
30. Greenwood, I. A., Miller, L. J., Ohya, S. & Horowitz, B. The large conductance potassium channel β -subunit can interact with and modulate the functional properties of a calcium-activated chloride channel, CLCA1. *J. Biol. Chem.* **277**, 22119–22122 (2002).
31. Qu, Z., Fischmeister, R. & Hartzell, C. Mouse bestrophin-2 is a bona fide Cl^- channel: identification of a residue important in anion binding and conduction. *J. Gen. Physiol.* **123**, 327–340 (2004).
32. Hartzell, H. C., Qu, Z., Yu, K., Xiao, Q. & Chien, L. T. Molecular physiology of bestrophins: multifunctional membrane proteins linked to best disease and other retinopathies. *Physiol. Rev.* **88**, 639–672 (2008).
33. Arreola, J. *et al.* Secretion and cell volume regulation by salivary acinar cells from mice lacking expression of the Clcn3 Cl^- channel gene. *J. Physiol. (Lond.)* **545**, 207–216 (2002).
34. Rosenfeld, M. A. & Collins, F. S. Gene therapy for cystic fibrosis. *Chest* **109**, 241–252 (1996).
35. Schwiebert, E. M., Benos, D. J., Egan, M. E., Stutts, M. J. & Guggino, W. B. CFTR is a conductance regulator as well as a chloride channel. *Physiol. Rev.* **79**, S145–S166 (1999).
36. Grubb, B. R., Vick, R. N. & Boucher, R. C. Hyperabsorption of Na^+ and raised Ca^{2+} -mediated Cl^- secretion in nasal epithelia of CF mice. *Am. J. Physiol.* **266**, C1478–C1483 (1994).
37. Boucher, R. C. *et al.* Chloride secretory response of cystic fibrosis human airway epithelia. Preservation of calcium but not protein kinase C- and A-dependent mechanisms. *J. Clin. Invest.* **84**, 1424–1431 (1989).
38. Anderson, M. P. & Welsh, M. J. Calcium and cAMP activate different chloride channels in the apical membrane of normal and cystic fibrosis epithelia. *Proc. Natl Acad. Sci. USA* **88**, 6003–6007 (1991).
39. Knowles, M. R., Clarke, L. L. & Boucher, R. C. Activation by extracellular nucleotides of chloride secretion in the airway epithelia of patients with cystic fibrosis. *N. Engl. J. Med.* **325**, 533–538 (1991).
40. Huang, X., Godfrey, T. E., Gooding, W. E., McCarty, K. S. Jr & Gollin, S. M. Comprehensive genome and transcriptome analysis of the 11q13 amplicon in human oral cancer and synteny to the 7F5 amplicon in murine oral carcinoma. *Genes Chromosom. Cancer* **45**, 1058–1069 (2006).
41. Carles, A. *et al.* Head and neck squamous cell carcinoma transcriptome analysis by comprehensive validated differential display. *Oncogene* **25**, 1821–1831 (2006).
42. Espinosa, I. *et al.* A novel monoclonal antibody against DOG1 is a sensitive and specific marker for gastrointestinal stromal tumors. *Am. J. Surg. Pathol.* **32**, 210–218 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank R. MacKinnon and J. Wood for a review of the manuscript. We also thank B. Hille for his technical comment on ion permeability. This work was supported by Acceleration Research of MOEST/KOSEF of Korea. This work was also supported by the Brain Korea 21 Project and by the Wellcome Trust Foundation, UK (R.R.).

Author Contributions Y.D.Y. cloned the channel and its mutants and GPCRs. W.-S.S. worked on bioinformatics. Y.D.Y. recorded currents in oocytes. H.C., B.L. and M.H.T. recorded whole-cell currents. J.Y.K. recorded single-channel currents. J.L. and Y.C. worked on siRNA *in vivo*. B.-M.K. worked on western blot. R.R. measured Ca^{2+} and currents. S.P.P. measured electrolytes in saliva. Y.K.S. worked on immunohistochemistry. U.O. designed and supervised experiments, and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to U.O. (utoh@snu.ac.kr).

METHODS

Cloning of mouse ANO1. The full-length coding sequence of mouse *Ano1* was amplified from first-strand cDNAs by PCR using site-specific primers: forward primer 5'-CCGCTCGAGGCCACCATGAGGGTCCCCGAGAAG-3', reverse primer 5'-GGGGTACCTACAGCGCGTCCCCATGGTAC-3'. pEGFP-N1 contained the *GFP* gene in the vector. To express ANO1–GFP fusion protein, a stop codon was deleted from pEGFP-N1-mANO1 using the Muta-Direct site-directed mutagenesis kit (iNtRON Biotech) because pEGFP-N1-mANO1 has a stop codon between the *Ano1* and *GFP* coding sequences.

Cloning of GPCRs. Full-length cDNAs of human ET_AR (*EDNRA*), human P2Y₂R (*P2Y2R*) and rat H1R (*Hrh1*) were obtained by PCR from single-strand cDNAs of HEK 293T, HaCaT and rat dorsal-root ganglion cells, respectively, and inserted into pCDNA3.1 (Invitrogen), pEGFP-N1 (Clontech) and pXOON (a gift from T. Jespersen). Rat AT1R (*Agtr1a*) and human M1R (*CHRM1*) were donated by Y. S. Bae and C. H. Lee, respectively.

siRNA. We designed three siRNAs that target different regions of the *Ano1* transcript and were positioned at nucleotides 1602–1620, 1935–1953 and 2288–2306. The targeted three siRNA sequences were: *Ano1*_siRNA1 (coding sequence position: 1602–1620): sense 5'-GAGUCUAGAGAAGUCACU(dTdT)-3', antisense 5'-AGUGACUUCUAAGACUC(dTdT)-3'; *Ano1*_siRNA2 (coding sequence position: 1935–1953): sense 5'-GAACAAUCUCUUGCA GAUU(dTdT)-3', antisense 5'-AAUCUCGAAGAGAUUGUUC(dTdT)-3'; and *Ano1*_siRNA3 (coding sequence position: 2288–2306): sense 5'-GACCT CAACCTCGAACCTT(dTdT)-3', antisense 5'-AAGGUUCGAGGUUGAAG UC(dTdT)-3'. Negative control siRNA (scrambled siRNA): sense 5'-CCU ACGCCACCAAUUUCGU(dTdT)-3', antisense 5'-ACGAAUUGGUGGC GUAGG(dTdT)-3'. The *Ano1* siRNAs and negative-control siRNA (scrambled siRNA) were synthesized (Bioneer Co.).

For efficient target delivery to tissues *in vivo*, *Ano1* siRNAs (10 µg each) were injected with polyethylenimine (*in vivo* jetPEI, Polyplus-transfection Inc.). The siRNA–polyethylenimine complex (200 µl) was injected to the retro-orbital sinuses in the orbits of 7-week-old male ICR mice.

To knockdown endogenous *Xenopus laevis* *Ano1* (*xAno1*), endonuclease-digested siRNA (esiRNA) was prepared using a Silencer siRNA Construction Kit (Ambion). Long double-stranded RNAs corresponding to nucleotides 1205–1503 of *xAno1* (NM_001096548, gi: 148228306) were transcribed, purified and digested using RNase III and injected (1 ng per 50 nl water) to oocytes. Silencer Negative Control number 1 siRNA (19 bp scrambled sequence with 3' dT overhangs, Ambion) was used as a negative control. Gene silencing was confirmed by RT–PCR.

Salivary secretion test. To measure the saliva volumes, anaesthetized mice (pentobarbital, 50 mg kg⁻¹, i.p.) were placed on a slight slope with heads facing downwards to avoid swallowing. A piece of pre-weighed cotton ball (~10 mg) was inserted below each tongue, and every 5 min the saliva-soaked cotton balls were removed, weighed, and then spun to extract saliva. Saliva osmolarities were measured using a Micro-Osmometer (Model 210, Fiske Associates), and Na⁺, Cl⁻ and K⁺ concentrations in extracted saliva were determined using an Autobiochemistry Analyser (Hitach 7070). Mice that vomited after pilocarpine injection were excluded.

Antibody production. The polyclonal antibody to mouse ANO1 was raised in a rabbit using a synthetic peptide affinity-purified by protein A-Sepharose column corresponding to the cytosolic region of ANO1 between TM2 and TM3 (451–466 amino acid: KDHPRAEYEARVLEKS; AbFrontier Co.). Anti-mouse ANO1

polyclonal antibody was purified by protein A-Sepharose column chromatography followed by antigen peptide-specific affinity column chromatography. The specificity of purified antibody was confirmed by immunoblot.

Immunohistochemical staining of ANO1. Sections (4 µm thickness) of 10% formalin-fixed, paraffin-embedded tissues were placed on silanized glass slides, deparaffinized three times with xylene, rehydrated in an alcohol series, and immersed in 0.3% H₂O₂. Sections were incubated with 1:800 diluted rabbit anti-mouse ANO1 polyclonal antibody followed by incubation with dextran polymer conjugated with goat anti-rabbit antibody and horseradish peroxidase (Envision plus kit, DAKO). The chromogen was developed over 5 min with liquid 3,3'-diaminobenzidine. The sections were counter-stained with Meyer's haematoxylin. Controls with normal rabbit serum and with blocking peptide were processed in parallel.

For immunofluorescent staining, frozen sections in 20–40 µm thickness of mouse submandibular glands were mounted on slide glass, fixed with methanol for 15 min, washed, and incubated with rabbit ANO1 polyclonal antibody (diluted 1:400, in blocking solution) and goat AQP5 polyclonal antibody (diluted 1:200, Santa Cruz Biotechnology) or goat NKCC1 polyclonal antibody (diluted 1:200, Santa Cruz Biotechnology) overnight at 4 °C. After wash, samples were incubated with Alexa Fluor 594-conjugated donkey anti-rabbit IgG (Molecular Probe, Inc., diluted 1:200) or Alexa Fluor 488-conjugated chicken anti-goat IgG (Molecular Probe, Inc., diluted 1:200) as a secondary antibody for 1 h.

Current recordings. To form gigaseals, individual cell surfaces were touched with a ~3 MΩ Sylgard-coated glass pipette with gentle suction. The membrane in contact with the pipette was then ruptured by suction to form a whole cell. The junction potential was adjusted to 0 mV. For ion selectivity experiments, 140 mM NMDG-Cl in the control bath solution was replaced with equimolar NaCl, NaI, NaBr, NaNO₃, NaF or Na-gluconate. For calculation of cation permeability, the pipette solution contained 70 mM NaCl, 2 mM MgCl₂, 70 mM sucrose and 10 mM HEPES (pH 7.2) and the bath solution contained 210 mM NaCl, 2 mM MgCl₂ and 10 mM HEPES (pH 7.2). To determine current–voltage relationships, voltage ramps from –80 to +80 mV were delivered for 350–400 ms to whole cells. Whole-cell currents were amplified using an Axopatch 200B amplifier (Molecular Device) and stored in a personal computer after digitization by a Digidata 1440 (Molecular Device).

For single-channel current recordings, the pipette and control bath solution contained 140 mM NMDG-Cl, 2 mM MgCl₂ and 10 mM HEPES (pH 7.2). Currents were amplified, filtered at 2 kHz, and stored in a data recorder, and when required these data were exported in a sampling rate of 100 KHz to a personal computer for current amplitude analysis. To create amplitude histograms, pClamp software (version 6.0, Molecular Device) was used to analyse open and closed single-channel current events. The half-amplitude algorithm was used to determine an open event. The minimum duration of an open event was set at 0.1 ms.

Oocyte recording. pSDTF-mANO1 was linearized using the BamHI restriction enzyme and transcribed with SP6 RNA polymerase (Ambion) to yield complementary RNA. After defolliculation with collagenase, *Xenopus laevis* oocytes were injected with 50 ng of the complementary RNA of ANO1 in 50 nl of diethylpyrocabonate-treated water. Two to five days after injection, whole-cell currents were recorded using the two-electrode voltage-clamp technique. The bath solution containing (in mM) 96 NaCl, 1 MgCl₂, 5 HEPES, 2 KCl and 1.8 CaCl₂ adjusted at pH 7.5 was perfused at 2 ml min⁻¹.

LETTERS

Squeezing and entanglement in a Bose–Einstein condensate

J. Estève¹, C. Gross¹, A. Weller¹, S. Giovanazzi¹ & M. K. Oberthaler¹

Entanglement, a key feature of quantum mechanics, is a resource that allows the improvement of precision measurements beyond the conventional bound attainable by classical means¹. This results in the standard quantum limit, which is reached in today's best available sensors of various quantities such as time² and position^{3,4}. Many of these sensors are interferometers in which the standard quantum limit can be overcome by using quantum-entangled states (in particular spin squeezed states^{5,6}) at the two input ports. Bose–Einstein condensates of ultracold atoms are considered good candidates to provide such states involving a large number of particles. Here we demonstrate spin squeezed states suitable for atomic interferometry by splitting a condensate into a few parts using a lattice potential. Site-resolved detection of the atoms allows the measurement of the atom number difference and relative phase, which are conjugate variables. The observed fluctuations imply entanglement between the particles^{7–9}, a resource that would allow a precision gain of 3.8 dB over the standard quantum limit for interferometric measurements.

Spin squeezing was one of the first quantum strategies proposed to overcome the standard quantum limit, in a precision measurement^{5,6} that triggered many experiments^{10–17}. It applies to measurements where the final readout is done by counting the occupancy difference

between two quantum states, as in interferometry or in spectroscopy. The name ‘spin squeezing’ originates from the fact that the N particles used in the measurement can be described by a fictitious spin $J = N/2$. In an interferometric sequence, the spin undergoes a series of rotations in which one of the rotation angles is the phase shift to be measured. A sufficient criterion for the input state, allowing for quantum-enhanced metrology, is given by $\xi_S < 1$, where $\xi_S^2 = 2J\Delta J_z^2 / (\langle J_x \rangle^2 + \langle J_y \rangle^2)$ is the squeezing parameter introduced in ref. 6. The fluctuations of the spin in one direction have to be reduced below shot noise (here $\Delta J_z^2 < J/2$), and the spin polarization in the orthogonal plane, $\langle J_x \rangle^2 + \langle J_y \rangle^2$, has to be large enough to maintain the sensitivity of the interferometer. A pictorial representation of this condition is shown in Fig. 1b. The precision of such a quantum-enhanced measurement is ξ_S/\sqrt{N} , whereas the standard quantum limit set by shot noise is $1/\sqrt{N}$.

In this Letter, we report the observation of entangled squeezed states in a Bose–Einstein condensate of ⁸⁷Rb atoms. The particles are distributed over a small number of lattice sites (between two and six) in a one-dimensional optical lattice (Fig. 1a). The occupation number per site ranges from 100 to 1,100 atoms. The two modes supporting the squeezing are two states of the external atomic motion corresponding to the condensate mean-field wavefunctions

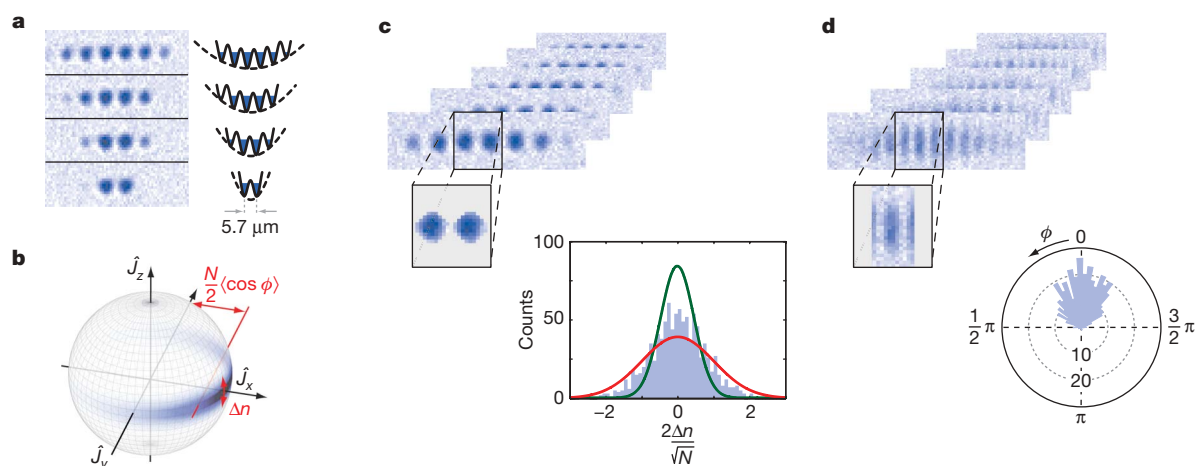


Figure 1 | Observing spin squeezing in a Bose–Einstein condensate confined in a double- or six-well trap. **a**, The atoms are trapped in an optical lattice potential superimposed on an harmonic dipole trap. The number of occupied sites is adjusted by changing the confinement in the lattice direction. High-resolution imaging allows us to resolve each site. **b**, Gain in quantum metrology is obtained for spin squeezed states exhibiting reduced fluctuations in one direction (z) and a sufficiently large polarization in the orthogonal plane (x, y) as depicted on the Bloch sphere. For our system, spin fluctuations in the z direction translate to atom number difference fluctuations Δn between two adjacent wells. The polarization of the spin in

the x – y plane is proportional to the phase coherence, $\langle \cos \phi \rangle$, between the wells. **c**, The atom number fluctuations at each site are measured by integrating the atomic density obtained from absorption images. We compare a typical histogram showing sub-Poissonian fluctuations in the atom number difference with the binomial distribution (red curve). The green curve corresponds to the deduced distribution after subtracting the photon shot noise, leading to a number squeezing factor of $\xi_N^2 = -6.6$ dB. **d**, The phase coherence is inferred from the interference patterns between adjacent wells. The histogram shown corresponds to a phase coherence of $\langle \cos \phi \rangle = 0.9$.

¹Kirchhoff-Institut für Physik, Universität Heidelberg, Im Neuenheimer Feld 227, 69120 Heidelberg, Germany.

in two adjacent lattice sites. These modes are spatially well separated and thus represent an ideal starting condition for a spatially split interferometer. The fictitious spin components can be defined as $J_x = (a^\dagger b + b^\dagger a)/2$, $J_y = i(a^\dagger b - b^\dagger a)/2$ and $J_z = (a^\dagger a - b^\dagger b)/2$, where a^\dagger and b^\dagger denote the respective creation operators associated with the two modes, a and b denote the corresponding annihilation operators, and $i = \sqrt{-1}$. The z component corresponds to half the atom number difference between the wells. Because the mean occupation numbers, n_a and n_b , in the two wells are large, the expectation values of the x and y components can be approximated by $\langle J_x \rangle \approx \sqrt{n_a n_b} \langle \cos \phi \rangle$ and $\langle J_y \rangle \approx \sqrt{n_a n_b} \langle \sin \phi \rangle$, respectively, where ϕ is the phase difference between the two macroscopic wavefunctions.

Spin squeezing by means of unitary evolution requires a nonlinear component in the Hamiltonian⁵; this is provided by the repulsive interactions between the atoms of the condensate. The corresponding suppression of atom number fluctuations in a Bose–Einstein condensate has been indirectly observed^{18–23}. However, by definition of the squeezing factor ξ_s , its experimental determination requires access to the local properties of the atoms occupying the two sites of interest. By imaging the condensate with a resolution of 1 μm (full-width at half-maximum), which is well below the lattice spacing of 5.7 μm , we fulfil this criterion of local measurement. The wells of the lattice are fully resolved (Fig. 1a), which allows for the determination of the atom number in each lattice site by direct integration of the atomic density as obtained by absorption imaging. Local interference measurements after a condensate expansion time short enough that only neighbouring sites overlap reveal the phase between these wells. In Fig. 1c, d we display typical data sets for the two types of measurement. The technical details of the precise experimental procedures and calculations used to deduce the number squeezing factor and the phase coherence are given in the Supplementary Information. The fluctuation measurement of the two conjugate variables, number and phase, yields information about the quantum state of the system and, in particular, allows for the detection of macroscopic entanglement between the particles.

Figure 2 summarizes the conjugate number–phase measurements in different experimental situations. The vertical axis corresponds to the number squeezing parameter, $\xi_N^2 = \Delta J_z^2 / \Delta J_{z,\text{ref}}^2$, which measures how much the fluctuations of the atom number difference are suppressed in comparison with a binomial distribution with variance $\Delta J_{z,\text{ref}}^2 = n_a n_b / N$, which is expected in a non-squeezed situation. The phase coherence, $\langle \cos \phi \rangle$, between the two wells defines the horizontal axis. We choose the origin of the phase such that the J_y component has a zero mean value. In this case, the relevant squeezing parameter for quantum metrology, ξ_s , is given by $\xi_s = \xi_N / \langle \cos \phi \rangle$. In the following, we will refer to it as the coherent number squeezing parameter. Lines corresponding to $\xi_s^2 = -3$ dB and $\xi_s^2 = -6$ dB are plotted in the figure. The different data points correspond to different preparations of the atoms in the lattice, as detailed in the legend of Fig. 3, and to different numbers of occupied sites. The simultaneous observation of number squeezing and high phase coherence (solid symbols) reveals the presence of coherent number squeezing. In the double-well and six-well situations, we deduce best squeezing factors of $\xi_s^2 = -2.3$ dB and $\xi_s^2 = -3.8$ dB, respectively. The statistical error bars and bounds for systematic errors are indicated in the figure. The open symbols show that atom number fluctuations can be further suppressed at the expense of lower phase coherence. In the six-well situation, we observe number squeezing down to $\xi_N^2 = -7.2$ dB, which corresponds to fluctuations of 15 atoms per well, out of 1,100. The inset in Fig. 2 shows the optimal number squeezing for a given phase coherence²⁴ in our experimental situation with 2,200 atoms in a double well, revealing that there is still great potential for improvement. Our best measurements yield number fluctuations approximately 25 dB higher than the Heisenberg limited states with the same phase coherence.

Entanglement in the context of spin squeezing has been intensively discussed^{17–9}. In a first quantization approach, it can be defined as the

nonseparability of the N -body density matrix. With this definition, a sufficient criterion for entanglement coincides with the criterion for quantum metrology, namely $\xi_s < 1$, which identifies spin-squeezing-type entanglement as a useful resource⁷. In the context of indistinguishable bosons, as in our experiment, the number squeezing $\xi_N < 1$ has been shown to be a sufficient criterion for the nonseparability of the reduced two-body density matrix^{8,9}. This criterion is satisfied for all the measurements shown in Fig. 2. However, without the possibility of accessing the single-particle spin properties (by contrast to the case of an ion string²⁵), this type of bipartite entanglement may not be usable as a resource. For this reason, only measurements in the coherent-number-squeezed region (solid symbols) indicate the definite presence of useful entanglement.

To identify what limits the amount of squeezing, we consider the two-mode Josephson Hamiltonian $E_C J_z^2 / 2 - 2E_J J_x / N$, that describes two weakly coupled condensates. The Josephson energy, E_J , and the charging energy, E_C , respectively characterize the tunnelling rate between the two condensates and the repulsive interaction energy inside each well. Because the depletion of the condensates in each well is small (≤ 12 atoms) we neglect intrawell excitation. Longer-wavelength excitations that exist in a many-well situation are also not considered. For the purpose of identifying the limiting factors on squeezing, this simplified model suffices because it captures the correct scaling of next-neighbour fluctuations with temperature, tunnelling rate and interaction energy²⁶. In the two-mode model and at temperatures T high in comparison with the plasma energy, $\sqrt{E_C E_J} / k_B$, the thermal excitation of the Josephson plasma mode limits the

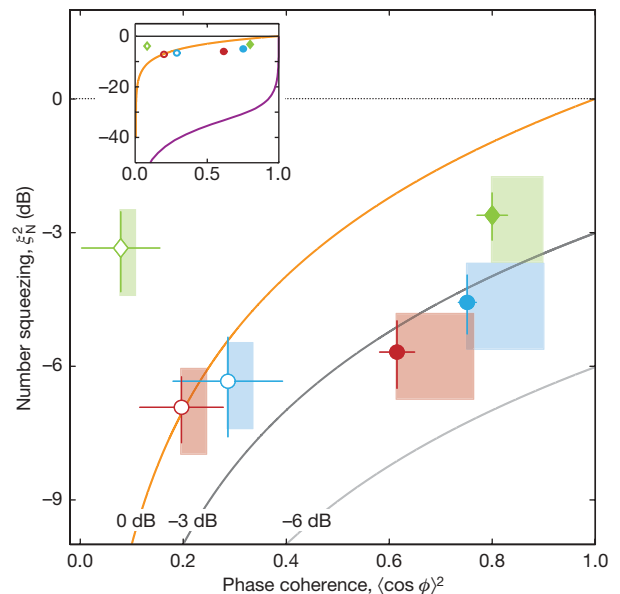


Figure 2 | Number squeezing and phase coherence as a criterion for quantum metrology and entanglement. Measurements are shown for the two main well pairs of a six-well lattice (red and blue circles) and for a double-well potential (green diamonds). The total atom number, N , in each pair is approximately 2,200 in the six-well case and 1,600 in the double-well case. Filled and open symbols respectively represent two different experimental regimes corresponding to different final lattice depths (see Fig. 3). For filled symbols, number squeezing and high phase coherence are simultaneously observed, whereas for the open symbols the phase coherence is degraded. The limits for different quantum metrology gains ξ_s^2 are plotted. In the coherent-number-squeezed region (below the orange line), directly usable entanglement is necessarily present in the system. The shaded areas show systematic error bounds due to a possible miscalibration of the atom number ($\pm 20\%$) and to an underestimation of the phase coherence caused by technical noise. The error bars indicate standard deviation deduced from at least 400 experimental realizations. The inset (same quantities as main panel) shows that our data are approximately 25 dB above the optimal allowed number squeezing (purple line), showing room for improvement²⁴.

number squeezing to $\xi_N^2 \approx k_B T / \mu$, where μ is the chemical potential that measures the strength of the interaction ($E_C \approx \mu / N$) and k_B is Boltzmann's constant. In a typical experimental situation at thermal equilibrium, these fluctuations place a strong limit on both number and coherent number squeezing, as it is difficult to evaporatively cool the condensate much below the chemical potential.

To circumvent this limitation, we prepare the condensate in a shallow lattice before increasing the lattice depth. During this splitting process, the tunnelling rate, E_J , decreases from its initial value, $E_J^{(i)}$, to a final value, $E_J^{(f)}$. The interaction energy, E_C , stays almost constant in our experimental situation. For sufficiently slow ramp speeds, the evolution is expected to be adiabatic. Because the energy of the collective Josephson mode decreases during the splitting, energy is removed from the system, leading to an effective cooling of the relevant degree of freedom. In the regime where only the linear part of the collective spectrum is populated, the effective temperature evolves as $T_{\text{eff}} = T \sqrt{E_J^{(f)} / E_J^{(i)}}$ leading to a final number squeezing $\xi_{N,f}^2 \approx \xi_{N,i}^2 (T_{\text{eff}} / T)$, where $\xi_{N,i}^2 = 4k_B T / N E_C$ is the initial squeezing at equilibrium with temperature T . The optimal number squeezing, assuming adiabatic evolution, is obtained by reducing further the Josephson energy to the Fock regime where the collective spectrum is quadratic. The final number squeezing is then given by $\xi_{N,f}^2 \approx \xi_{N,i}^2 (k_B T / 2E_J^{(i)})$.

We investigate our assumption of adiabaticity by performing a series of experiments in the six-well situation. The lattice depth is ramped from the initial value at which the condensate is obtained (430 Hz) to a fixed end value (1,650 Hz) in different times. In Fig. 3a, the final number squeezing factors for the two most populated well pairs are plotted versus the total ramping time. As expected, no improvement in the number squeezing is observed for ramps that occur too quickly (< 20 ms). For slower ramps, the number squeezing increases and saturates at $-6.6_{-1.0-0.8}^{+0.8+0.8}$ dB, where the given uncertainties are

1σ statistical errors obtained from approximately 1,000 experimental realizations followed by bounding of the systematic error (see Supplementary Information). This improvement in number squeezing by a factor of approximately three is a consequence of the effective cooling discussed above. As shown in Fig. 3a, the observed behaviour is reproduced by a numerical simulation of the splitting process with the two-mode Josephson Hamiltonian.

High phase coherence is the second ingredient for a coherent number squeezing factor $\xi_S < 1$. This requires that the effective temperature of the Josephson mode be below the tunnelling coupling E_J / k_B , which is the case for the initial lattice depth²⁷. Increasing the lattice depth decreases the phase coherence. Best coherent number squeezing is obtained by adiabatically increasing the lattice depth to an optimal value such that coherence is still high and number squeezing has occurred. In order to find this optimum, we investigate the number squeezing and the phase coherence as a function of tunnelling coupling. We linearly ramp up the barrier height to different end values, keeping the ramp speed in the adiabatic regime. Figure 3 presents the results obtained for six (Fig. 3b) and two (Fig. 3c) occupied wells. In the six-well case, we identify an optimum barrier height range between 650 and 900 Hz, where we deduce a best coherent number squeezing $\xi_S^2 = -3.8_{-0.4-1.8}^{+0.3+0.8}$ dB averaging over all the measurements. In the double-well situation, averaging all the points between 650 and 1,200 Hz, we obtain $\xi_S^2 = -2.3_{-0.6-1.5}^{+0.2+0.8}$ dB. The deduced mean phase coherence and number squeezing in the optimal regions correspond to the solid data points shown in Fig. 2.

For the best observed number squeezing, we measure atom number fluctuations just above the detection threshold of our absorption imaging technique. To reach this sensitivity level, special care has to be taken to calibrate the deduced atom number²⁸ (see Supplementary Information). Furthermore, the contribution of the photon shot noise has to be precisely measured and subtracted²⁹. As an independent check of the reliability of the atom counting, we monitor the evolution

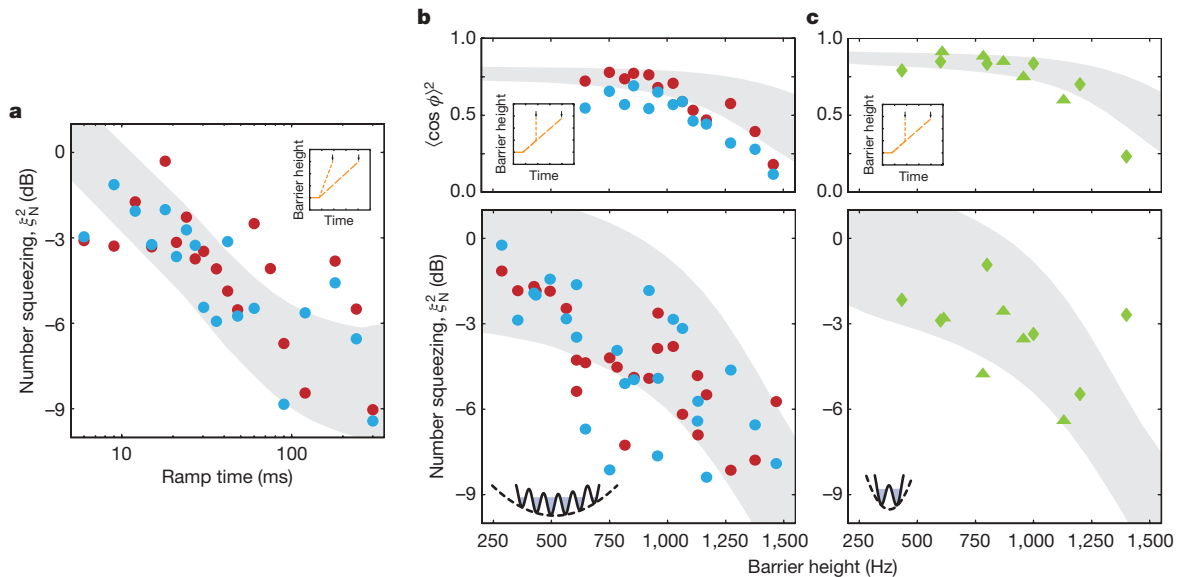


Figure 3 | Systematics of number squeezing and phase coherence during the splitting of a condensate. **a**, Investigating the adiabaticity condition for the evolution of the many-body state during the splitting process. The number squeezing factors for the two most populated well pairs (blue and red circles) are measured after ramping up the lattice depth from the value at which the condensate is obtained (430 Hz) to a fixed end value (1,650 Hz), for a number of different total ramp times (see inset). Numerical simulations using the two-mode Josephson model reproduce the general observed behaviour, assuming an initial thermal population of the density matrix corresponding to temperatures between 20 nK and 40 nK (grey shaded area). **b**, Number squeezing and phase coherence for different final lattice depths in the six-well case. The ramp speed is fixed to 4 Hz ms^{-1} (300-ms ramp time in **a**) to satisfy the adiabaticity criterion. Number squeezing improves before

the phase coherence drops, leading to coherent number squeezing. The grey shaded area shows the predictions of the two-mode model assuming an adiabatic evolution of the density matrix with initial temperatures between 10 nK and 30 nK (three to ten populated many-body states). **c**, Same as **b** but for a double well. The ramp speeds are 2 Hz ms^{-1} and 8 Hz ms^{-1} for the triangles and diamonds, respectively. The symbols shown in Fig. 2 correspond to the averages of all the data measured for barrier heights between 650 and 900 Hz (filled symbols) and above 1,300 Hz (open symbols) in the six-well case (**b**) and between 650 and 1,200 Hz (filled symbols) and above 1,400 Hz (open symbols) in the double-well case (**c**). Some data measured at very high lattice depth in the double-well case that participate in the averaging are not shown here.

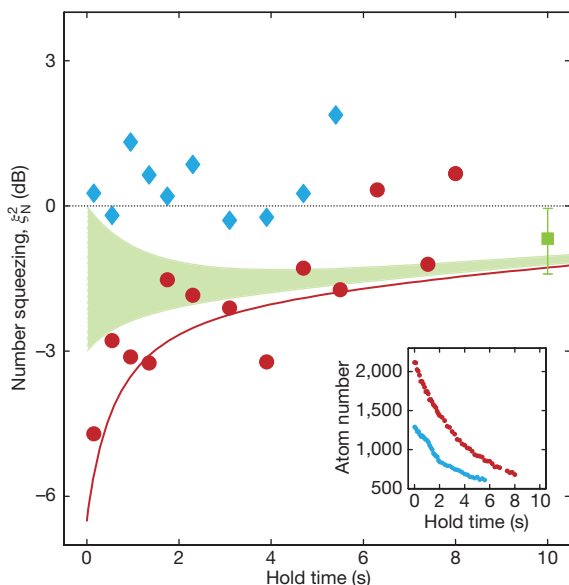


Figure 4 | Random loss restores Poissonian fluctuations. With the atoms held in the lattice (six occupied sites), the atom number in the decoupled two main well pairs decreases as shown in the inset, owing to one- and three-body loss. Red circles show the evolution of number squeezing starting from a squeezed situation. The solid line is the prediction corresponding to the measured loss rates and initial squeezing. Blue diamonds show that in the absence of initial squeezing no further evolution is observed. The consistent behaviour of the squeezing with loss is an independent check of the validity of our squeezing measurement technique. A more quantitative test is performed by holding the atoms for 10 s and measuring the number squeezing inferred from 1,000 realizations. The initial state is slightly squeezed ($-3 \text{ dB} < \xi_N^2 < 0 \text{ dB}$) and a squeezing factor $-1.2 \text{ dB} < \xi_N^2 < -1 \text{ dB}$ (green shaded area) is expected at the time of measurement, when two-thirds of the atoms have been lost. We measure $\xi_N^2 = -0.7^{+0.7}_{-0.7} \text{ dB}$, which is consistent with the expected value within the 95% statistical confidence bounds indicated by the error bar. The small improvement of number squeezing in the upper boundary of the green shaded area arises from the nonlinear dependence of three-body loss on atom number. After a sufficient time, the one-body loss dominates and Poissonian fluctuations are restored.

of number squeezing with atom loss. It is well known in quantum optics that random loss processes rapidly degrade the number squeezing. Red circles in Fig. 4 show this restoring effect with a rate that is compatible with the measured one- and three-body loss rates. A further check is performed by monitoring the evolution of the number fluctuations for an initial state with almost Poissonian fluctuations; such a state is prepared by directly condensing the atoms in a deep lattice. No change in the number squeezing is observed through the loss process, within the statistical errors. During the measurement, the relevant parameters for the imaging, such as the extension of the cloud and its optical density, are kept constant.

To confirm the successful implementation of a quantitative atom number fluctuation measurement, we infer number squeezing from 1,000 measurements after a 10-s holding time, starting from a slightly squeezed situation ($-3 \text{ dB} < \xi_N^2 < 0 \text{ dB}$). At this time, two-thirds of the atoms are lost and a precise number squeezing factor $-1.2 \text{ dB} < \xi_N^2 < -1 \text{ dB}$ is predicted from the measured one- and three-body loss coefficients. This exemplifies how particle losses can be used to prepare a well-defined number distribution in a condensate. We measure $\xi_N^2 = -0.7^{+0.7}_{-0.7} \text{ dB}$, where the indicated uncertainties are 95% statistical confidence bounds, in quantitative agreement with the expected number squeezing.

The measured squeezing presented here concerns the external degree of freedom of the atoms in the condensate. It is comparable in amount to the latest measured squeezing of internal atomic states, also obtained by unitary evolution of a nonlinear Hamiltonian^{13,14}.

We show that the achieved entanglement can be directly used as a resource for quantum metrology with spatial atom interferometers. This is a major step towards measuring at the ultimate, Heisenberg, limit with a large number of particles.

Received 2 May; accepted 12 August 2008.

Published online 1 October 2008; corrected 31 October 2008 (details online).

- Giovannetti, V., Lloyd, S. & Maccone, L. Quantum-enhanced measurements: Beating the standard quantum limit. *Science* **306**, 1330–1336 (2004).
- Santarelli, G. *et al.* Quantum projection noise in an atomic fountain: A high stability cesium frequency standard. *Phys. Rev. Lett.* **82**, 4619–4622 (1999).
- Goda, K. *et al.* A quantum-enhanced prototype gravitational-wave detector. *Nature Phys.* **4**, 472–476 (2008).
- Arcizet, O. *et al.* High-sensitivity optical monitoring of a micromechanical resonator with a quantum-limited optomechanical sensor. *Phys. Rev. Lett.* **97**, 133601 (2006).
- Kitagawa, M. & Ueda, M. Squeezed spin states. *Phys. Rev. A* **47**, 5138–5143 (1993).
- Wineland, D. J., Bollinger, J. J., Itano, W. M. & Heinzen, D. J. Squeezed atomic states and projection noise in spectroscopy. *Phys. Rev. A* **50**, 67–88 (1994).
- Sørensen, A., Duan, L., Cirac, J. & Zoller, P. Many-particle entanglement with Bose–Einstein condensates. *Nature* **409**, 63–66 (2001).
- Wang, X. & Sanders, B. C. Spin squeezing and pairwise entanglement for symmetric multiqubit states. *Phys. Rev. A* **68**, 012101 (2003).
- Korbicz, J. K., Cirac, J. I. & Lewenstein, M. Spin squeezing inequalities and entanglement of N qubit states. *Phys. Rev. Lett.* **95**, 120502 (2005).
- Hald, J., Sørensen, J. L., Schori, C. & Polzik, E. S. Spin squeezed atoms: A macroscopic entangled ensemble created by light. *Phys. Rev. Lett.* **83**, 1319–1322 (1999).
- Kuzmich, A., Mandel, L. & Bigelow, N. P. Generation of spin squeezing via continuous quantum nondemolition measurement. *Phys. Rev. Lett.* **85**, 1594–1597 (2000).
- Julsgaard, B., Kozhekin, A. & Polzik, E. S. Experimental long-lived entanglement of two macroscopic objects. *Nature* **413**, 400–403 (2001).
- Chaudhury, S. *et al.* Quantum control of the hyperfine spin of a Cs atom ensemble. *Phys. Rev. Lett.* **99**, 163002 (2007).
- Fernholz, T. *et al.* Spin squeezing of atomic ensembles via nuclear-electronic spin entanglement. Preprint at (<http://arxiv.org/abs/0802.2876>) (2008).
- Meyer, V. *et al.* Experimental demonstration of entanglement-enhanced rotation angle estimation using trapped ions. *Phys. Rev. Lett.* **86**, 5870–5873 (2001).
- Leibfried, D. *et al.* Toward Heisenberg-limited spectroscopy with multiparticle entangled states. *Science* **304**, 1476–1478 (2004).
- Roos, C. F. *et al.* Control and measurement of three-qubit entangled states. *Science* **304**, 1478–1480 (2004).
- Orzel, C., Tuchman, A., Fenselau, M., Yasuda, M. & Kasevich, M. Squeezed states in a Bose–Einstein condensate. *Science* **291**, 2386–2389 (2001).
- Greiner, M., Mandel, O., Hansch, T. & Bloch, I. Collapse and revival of the matter wave field of a Bose–Einstein condensate. *Nature* **419**, 51–54 (2002).
- Gerbier, F., Fölling, S., Widera, A., Mandel, O. & Bloch, I. Probing number squeezing of ultracold atoms across the superfluid–Mott insulator transition. *Phys. Rev. Lett.* **96**, 090401 (2006).
- Sebby-Strabley, J. *et al.* Preparing and probing atomic number states with an atom interferometer. *Phys. Rev. Lett.* **98**, 200405 (2007).
- Jo, G.-B. *et al.* Long phase coherence time and number squeezing of two Bose–Einstein condensates on an atom chip. *Phys. Rev. Lett.* **98**, 030407 (2007).
- Li, W., Tuchman, A. K., Chien, H.-C. & Kasevich, M. A. Extended coherence time with atom-number-squeezed states. *Phys. Rev. Lett.* **98**, 040402 (2007).
- Sørensen, A. S. & Mølmer, K. Entanglement and extreme spin squeezing. *Phys. Rev. Lett.* **86**, 4431–4434 (2001).
- Korbicz, J. K. *et al.* Generalized spin-squeezing inequalities in N -qubit systems: Theory and experiment. *Phys. Rev. A* **74**, 052319 (2006).
- Javanainen, J. Phonon approach to an array of traps containing Bose–Einstein condensates. *Phys. Rev. A* **60**, 4902–4909 (1999).
- Gati, R., Hemmerling, B., Fölling, J., Albiez, M. & Oberthaler, M. K. Noise thermometry with two weakly coupled Bose–Einstein condensates. *Phys. Rev. Lett.* **96**, 130404 (2006).
- Reinaudi, G., Lahaye, T., Wang, Z. & Guéry-Odelin, D. Strong saturation absorption imaging of dense clouds of ultracold atoms. *Opt. Lett.* **32**, 3143–3145 (2007).
- Estève, J. *et al.* Experimental observations of density fluctuations in an elongated Bose gas: Ideal gas and quasi-condensate regimes. *Phys. Rev. Lett.* **96**, 130403 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We gratefully acknowledge support from the DFG, GIF and EC (MIDAS STREP). J.E. acknowledges support from the EC Marie-Curie program. C.G. acknowledges support from the Landesgraduiertenförderung Baden-Württemberg.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to M.K.O. (entanglement@matterwave.de).

LETTERS

Marginal breakdown of the Fermi-liquid state on the border of metallic ferromagnetism

R. P. Smith¹, M. Sutherland¹, G. G. Lonzarich¹, S. S. Saxena¹, N. Kimura², S. Takashima³, M. Nohara³ & H. Takagi^{3,4}

For the past half century, our understanding of how the interactions between electrons affect the low-temperature properties of metals has been based on the Landau theory of a Fermi liquid¹. In recent times, however, there have been an increasingly large number of examples in which the predictions of the Fermi-liquid theory appear to be violated². Although the qualitative reasons for the breakdown are generally understood, the specific quantum states that replace the Fermi liquid remain in many cases unclear. Here we describe an example of such a breakdown where the non-Fermi-liquid properties can be interpreted. We show that the thermal and electrical resistivities in high-purity samples of the *d*-electron metal ZrZn₂ at low temperatures have T and $T^{5/3}$ temperature dependences, respectively: these are the signatures of the ‘marginal’ Fermi-liquid state^{3–7}, expected to arise from effective long-range spin–spin interactions in a metal on the border of metallic ferromagnetism in three dimensions^{3,5}. The marginal Fermi liquid provides a link between the conventional Fermi liquid and more exotic non-Fermi-liquid states that are of growing interest in condensed matter physics. The idea of a marginal Fermi liquid has also arisen in other contexts—for example, in the phenomenology of the normal state of the copper oxide superconductors⁷, and in studies of relativistic plasmas and of nuclear matter^{3,4,6}.

The low-temperature properties of metals are traditionally described in terms of the concept of elementary excitations. These excitations include the quanta of motion of the crystal lattice, or phonons, and the quanta of motion of the conduction electron system, or quasiparticle–quasihole pairs. The density of elementary excitations in thermal equilibrium decreases with decreasing temperature and vanishes as $T \rightarrow 0$ K. If the interactions between these excitations are short range, we then expect the characteristic lifetime between collisions, τ , to diverge with decreasing temperature: that is, we expect the scattering rate, τ^{-1} , to vanish as $T \rightarrow 0$. A key property of the Fermi-liquid state is that the quasiparticle scattering rate vanishes specifically as the square of the temperature, that is, $\tau^{-1} \propto T^2$. This leads to a T^2 temperature dependence of both the thermal resistivity, w , and the electrical resistivity, ρ , a behaviour characteristic of conventional metals at low temperatures.

The Fermi-liquid state can break down if the quasiparticle interactions are attractive or if they are long range. In particular, the conventional (relativistic) current–current interaction between charge carriers remains long range in a normal metal and thus could, at least in principle, lead to a non-Fermi-liquid form of the relaxation rate τ^{-1} at sufficiently low temperatures. For an ideally pure system in three dimensions (3D), the current–current interaction is expected to give rise to a quasiparticle scattering rate of the anomalous form $\tau^{-1} \propto T \log(T^*/T)$ in the $T \rightarrow 0$ limit, where T^* is a temperature scale⁶. This leads to a linear temperature dependence of the thermal resistivity, $w \propto T$, and a five-thirds power law for the electrical

resistivity, $\rho \propto T^{5/3}$ (refs 4–6). This is a crucial difference from the Fermi-liquid state, as it means that the characteristic uncertainty in the energy of a thermally excited quasiparticle (proportional to τ^{-1}) diverges in comparison with its characteristic energy (proportional to T). This effectively means that there are no well-defined fermionic excitations and that there is, evidently, a breakdown not only of the Fermi-liquid state but also potentially of the concept of elementary excitations itself, as it is traditionally understood. The state in which the ratio τ^{-1}/T diverges logarithmically is the marginal case in which the quasiparticles cease to exist.

This is an example of a type of marginal Fermi liquid that arises as a result of long-range (non-local) effective quasiparticle interactions^{3–6}. The term ‘marginal Fermi liquid’ was first introduced to describe a different problem, involving local interactions in a model of the high-temperature copper oxide superconductors⁷. In both local and non-local marginal Fermi liquids, the thermal resistivity, w , is linear in T at low T . However, crucially, the electrical resistivities differ in the two cases, namely, ρ varies as T and as $T^{5/3}$ in the local and above non-local marginal Fermi liquids, respectively. We shall be concerned here solely with the non-local case in which the T dependences of w and ρ differ such that the ratio w/ρ does not saturate at low T , but diverges as $T^{-2/3}$.

The current–current interaction is important in relativistic plasmas, but can usually be ignored for normal metallic densities under achievable experimental conditions. A marginal Fermi-liquid state might arise at readily accessible conditions, on the other hand, owing to the effects of long-range spin–spin interactions in pure metals on the border of ferromagnetic order at low temperatures. The problem of long-range spin–spin interactions has been described by the self-consistent-renormalization (SCR) approximation⁸ and related approximations^{3,4,9}, and also by the renormalization group techniques applied to quantum phase transitions and quantum critical phenomena¹⁰. The predictions of these approaches for the transport properties are qualitatively similar, and are summarized in Table 1 for a 3D itinerant-electron system on the border of ferromagnetism.

The SCR model reduces to the Fermi-liquid (FL) model below a temperature scale T_{FL} , which tends to vanish at the critical pressure or critical composition where ferromagnetic order disappears. Above T_{FL} but below another crossover temperature T_{MFL} (to be discussed below in a particular case), the SCR model reduces to the marginal Fermi liquid (MFL) if the ferromagnetic transition is continuous^{8,9}. The marginal Fermi-liquid regime has been difficult to identify unambiguously because (1) the transitions of interest are typically first order rather than continuous^{11,12}, (2) phonons can mask the intrinsic electronic state at intermediate temperatures (that is, $T_{\text{FL}} < T < T_{\text{MFL}}$) and (3) quenched disorder not included in the SCR model can play an important role in some of the key materials investigated thus far^{13,14}.

¹Cavendish Laboratory, University of Cambridge, J. J. Thomson Avenue, Cambridge CB3 0HE, UK. ²Center for Low Temperature Science, Tohoku University, Sendai, Miyagi 980-8578, Japan. ³Department of Advanced Materials Science, University of Tokyo, Kashiwa, Chiba 277-8561, Tokyo, Japan. ⁴RIKEN, 2-1 Hirosawa Wako, Saitama 351-0198, Japan.

Table 1 | Temperature dependences of transport properties

Transport property	Fermi liquid	Marginal Fermi liquid
Electrical resistivity	$\rho = \rho_0 + AT^2$	$\rho = \rho_0 + aT^{5/3}$
Thermal resistivity, $w = L_0T/\kappa$	$w = w_0 + AT^2 + BT^2$	$w = w_0 + aT^{5/3} + bT$
Thermal minus electrical resistivities, $\delta = w - \rho$	$\delta = BT^2$	$\delta = bT$

Results are shown for a Fermi liquid and for a marginal Fermi liquid with long-range spin–spin interactions on the border of ferromagnetism^{3,5,33}. For a comparison of the temperature dependences of other physical properties, see, for example, refs 8, 9. The specific heat for the marginal Fermi liquid differs from that of a Fermi liquid by a correction factor that is logarithmic in T . In contrast to w and ρ , the correction is weak and difficult to separate from phonon and other contributions to the specific heat in materials such as ZrZn_2 .

In this paper we consider an example in which these difficulties have been overcome, and in which the signatures of the marginal Fermi-liquid state in the transport properties (w and ρ) have been observed and interpreted in terms of a semiquantitative model for the first time.

The material selected for the search for the marginal Fermi-liquid state is ZrZn_2 , which crystallizes in the C15 cubic Laves structure. This compound is the best-known member of a small group of marginally ferromagnetic metals in which the constituent elemental metals are not magnetic (in fact, Zr and Zn are superconductors at low temperature). The low-temperature magnetization M and the Curie temperature T_C of ZrZn_2 are both an order of magnitude smaller than that of nickel, the archetypal itinerant-electron ferromagnet, and the magnetic transition is continuous except very close to the critical pressure where ferromagnetic order disappears¹⁵. Also, Fermi surface measurements¹⁶ show that the ferromagnetism in ZrZn_2 can be understood in terms of itinerant electrons alone, and inelastic neutron scattering studies^{17,18} reveal that the spin-fluctuation spectrum is principally characterized by dissipative modes of a form expected in the SCR model.

Importantly, ZrZn_2 can be purified to a sufficiently high level to allow us to unambiguously separate out the electronic and phonon contributions to w . This has thus far been the central stumbling block in establishing the crucial difference between w and ρ expected for a

marginal Fermi-liquid state in a nearly or weakly ferromagnetic metal. The details of the measurements of w and ρ , the separation of the electronic and phonon contributions and numerical calculations of w and ρ in the SCR model are given in the Supplementary Information.

The ambient pressure measurements of the transport properties in a high purity sample and comparisons with the predictions of the SCR model are presented in Fig. 1. Figure 1a and b shows the temperature variations of the thermal conductivity, κ , and of ρ , respectively, for a sample with a residual resistivity of $\rho_0 = 0.31 \mu\Omega \text{ cm}$. (This corresponds to a residual resistivity ratio $\rho(373 \text{ K})/\rho_0$ of close to 200 and an electronic mean free path of several thousand ångströms.) The inset of Fig. 1a shows κ versus T for this sample and for a less clean sample over a wide temperature range. A comparison of these two samples (Supplementary Information, section 3) allows us to conclude that below 15 K the thermal conductivity of the high-quality sample is overwhelmingly dominated by electrons. Figure 1c shows the difference, δ , between the thermal resistivity, which we define using the Wiedemann–Franz law to be $w = L_0T/\kappa$, and ρ , plotted versus temperature. Here L_0 is the Sommerfeld value of the Lorenz number ($2.45 \times 10^{-8} \text{ W } \Omega \text{ K}^{-2}$). Figure 1d shows the temperature-dependent part of the resistivity, $\Delta\rho = \rho - \rho_0$, plotted in this case versus $T^{5/3}$.

We see that δ is linear in T and $\Delta\rho$ is linear in $T^{5/3}$ at low temperatures. As shown in the insets of Fig. 1c and d, the observed

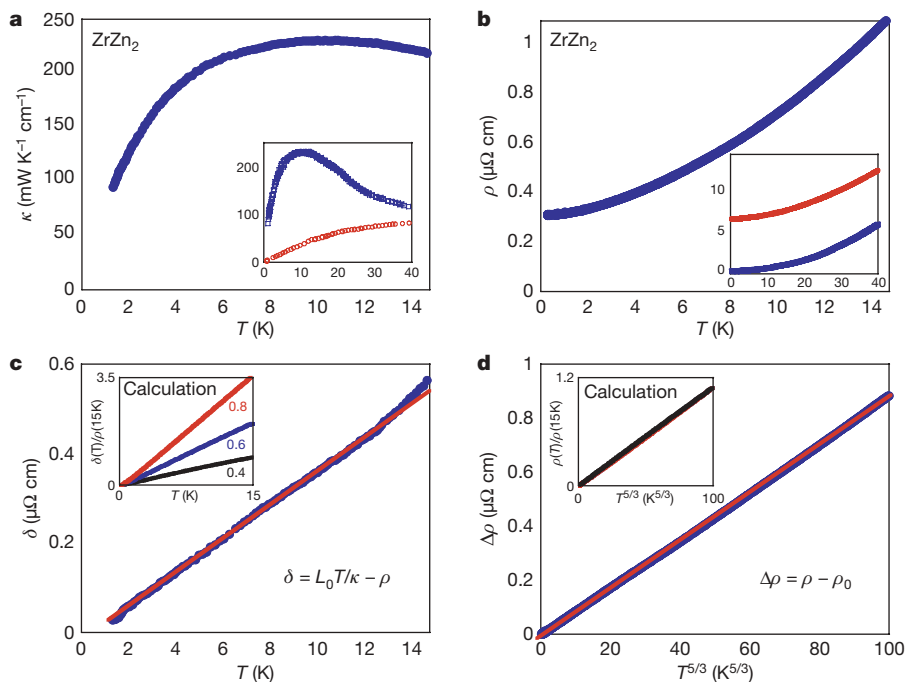


Figure 1 | Evidence for the marginal Fermi-liquid state in ZrZn_2 . **a, b,** Main panels, the thermal conductivity (**a**) and electrical resistivity (**b**) of ZrZn_2 (blue curves; residual resistivity $\rho_0 = 0.31 \mu\Omega \text{ cm}$). Insets in **a** and **b** show the same properties over a wider temperature range, and also for a sample with higher ρ_0 (red curves; $6.9 \mu\Omega \text{ cm}$). In this figure and Fig. 3 the error for each point is believed to be smaller than the size of the data points. Comparison of these two samples allows us to demonstrate that the thermal conductivity of the clean sample is strongly dominated by conduction by electrons up to 15 K (Supplementary Information, section 3). **c,** The difference between the

thermal and the electrical resistivity (δ) is linear with temperature, as predicted for the marginal Fermi liquid. **d,** The temperature-dependent part of the electrical resistivity plotted against $T^{5/3}$ also shows the marginal Fermi-liquid dependence (the straight lines in **c** and **d** are guides to the eye). Insets in **c** and **d** show the predictions of the SCR model for ZrZn_2 for three different values of the parameter k_c (in \AA^{-1}); in both cases the values are scaled to the electrical resistivity at 15 K. Details of these calculations are described in Supplementary Information, section 4.

behaviour is consistent qualitatively and even semiquantitatively with the predictions of the SCR model for ZrZn_2 (see also Supplementary Information, section 4). The predictions of the model are relatively insensitive to the precise model parameters. The most sensitive parameter is the characteristic diameter, k_c , of the Fermi surface of the dominant charge carriers, and the value that gives the best quantitative agreement is in keeping with the known Fermi surface of ZrZn_2 (Fig. 1c inset).

The SCR model also predicts the orders of magnitude of T_{FL} , T_{MFL} and T_C . The calculations of these quantities versus normalized pressure (p/p_c , where p_c is the critical pressure where ferromagnetism is suppressed completely) for ZrZn_2 are shown in Fig. 2. Just below p_c , where T_{MFL} is small and the phonon contribution to transport can be ignored in the marginal Fermi-liquid regime, we find a $T^{5/3}$ variation of the electrical resistivity both below and above T_C , with an abrupt change in slope at T_C (Fig. 3). As shown in Fig. 3 inset, this behaviour is again consistent with the prediction of the SCR model. We observe an electrical resistivity exponent of approximately $5/3$ at low T over the entire pressure range up to p_c (Fig. 4). The sudden drop in the exponent at higher pressures (where the magnetic transition becomes first order¹⁵) and beyond ($p \geq p_c$) will be discussed below.

Our findings thus provide strong evidence for the existence of marginal Fermi-liquid behaviour in ZrZn_2 . The new results are that (1) the thermal resistivity is linear in temperature and that this behaviour is consistent with a $T^{5/3}$ temperature dependence of the resistivity within the marginal Fermi-liquid model, and (2) that the temperature dependences of the thermal and electrical resistivities are in agreement with the predictions of the SCR theory. This theory reduces to the theory of a marginal Fermi liquid over large portions of the temperature–pressure phase diagram of ZrZn_2 . The signature of the marginal Fermi-liquid model should also appear in the temperature dependence of the specific heat and of the magnetic equation of state. The marginal Fermi-liquid form of the specific heat differs from that of a Fermi liquid by a factor logarithmic in T , which is weak and difficult to separate reliably from the phonon contribution in ZrZn_2 in the regime of interest, $T_{\text{FL}} < T < T_{\text{MFL}}$ (ref. 19; see also Supplementary Information). This difficulty extends to a number of other d -electron systems, including the optimally doped copper oxides in which the predicted logarithmic correction to the specific heat is obscured by phonons and other effects⁷. The signature of the marginal Fermi-liquid state has also been difficult to identify in the magnetic equation of state for two reasons. These are, first, the extreme sensitivity of the uniform magnetization to magnetic inhomogeneities and, second, the complex form of the free energy as a

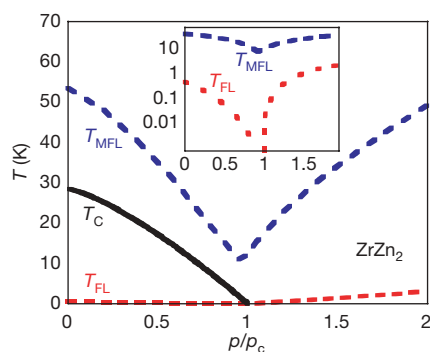


Figure 2 | Predicted phase diagram of ZrZn_2 . Main panel, the solid line represents the Curie temperature T_C , and the two dashed lines correspond to crossover temperatures. Below the lower dashed line (T_{FL}) the SCR model reduces to the Fermi-liquid model, whereas between the lower dashed line and upper dashed line (T_{MFL}) the SCR model reduces essentially to a marginal Fermi-liquid (MFL) model (except in a narrow region near T_C). Inset, the same diagram with a logarithmic temperature scale. T_{FL} is defined by the resistivity temperature exponent dropping to 1.85, and T_{MFL} is defined as where the resistivity exponent drops to 1.55.

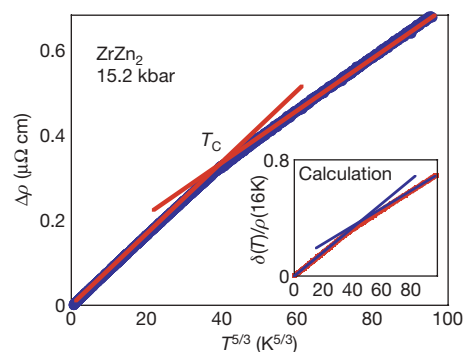


Figure 3 | Resistivity of ZrZn_2 just below the critical pressure. Main panel, the measured resistivity of ZrZn_2 at 15.2 kbar plotted against $T^{5/3}$, showing the $T^{5/3}$ behaviour of the resistivity both below and above T_C but with a different slope. Inset, the SCR model prediction, which shows excellent qualitative (and semiquantitative; see Supplementary Information) agreement. The straight lines in these figures are guides to the eye.

function of the uniform magnetization, which precludes a simple expansion of the magnetic equation of state as a power series in the magnetization^{15,20}. Crucially, these and other previously mentioned difficulties are absent or less severe for the case of the transport properties, which have therefore been the focus of our attention.

We now consider other evidence of marginal Fermi-liquid behaviour in weakly ferromagnetic metals. First, a $5/3$ power law for the temperature dependence of the electrical resistivity is not rare in such systems, and has been seen recently not only in ZrZn_2 (refs 19, 21) but also, for example, in $\text{Pd}_{1-x}\text{Ni}_x$ (ref. 13) and Ni_3Al (ref. 22). Second, a logarithmic correction to the specific heat has been identified in $\text{Pd}_{1-x}\text{Ni}_x$ (ref. 13) and NbFe_2 (ref. 23). (Such a logarithmic correction is also ubiquitous in f -electron metals near various subtle kinds of quantum critical points².) The observation of both a $T^{5/3}$ electrical resistivity and $T \ln(T^*/T)$ specific heat in $\text{Pd}_{1-x}\text{Ni}_x$ in particular, is compelling evidence for the existence of the marginal Fermi-liquid state of the non-local kind we are considering here. However, because $\text{Pd}_{1-x}\text{Ni}_x$ is intrinsically disordered, a clear separation of the electronic and phonon contributions to the thermal resistivity that would help to confirm the marginal Fermi-liquid scenario would seem to be ruled out in this system. We also note that a quantitative

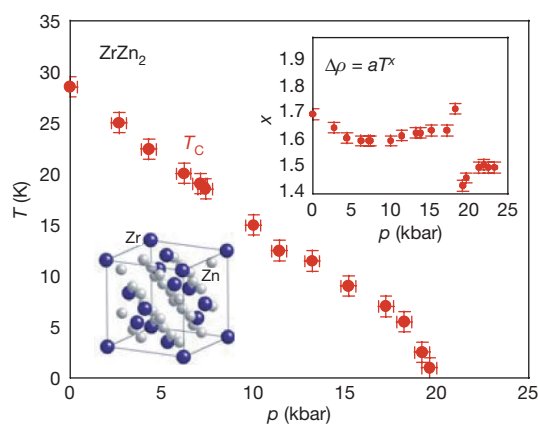


Figure 4 | Pressure dependence of T_C and the resistivity exponent in ZrZn_2 . Main panel, the pressure dependence of the Curie temperature T_C as inferred from the kink in the resistivity. Error bars on pressure show the width of the superconducting transition of tin used to determine the pressure; error bars on temperature show the difference between several methods of determining the temperature of the kink in the resistivity. Upper inset, the temperature exponent of the resistivity (in the range 1–5 K) versus pressure, showing the abrupt change from around $5/3$ to $3/2$ at around 18 kbar. Error bars on x are determined from the uncertainty in the power-law fit to the data. Lower inset, the cubic Laves structure of ZrZn_2 .

description of $\text{Pd}_{1-x}\text{Ni}_x$ along the lines given here for ZrZn_2 has not yet been carried out to our knowledge, and may not be possible within the SCR model in its present form¹⁴. Despite these difficulties, the findings in $\text{Pd}_{1-x}\text{Ni}_x$ and other low- T_C ferromagnetic metals in general provide further support for the idea that the marginal Fermi-liquid state can exist over at least some portion of the temperature–pressure phase diagram in the weakly ferromagnetic regime ($p \lesssim p_c$), when the magnetic transition is continuous or not too strongly first order.

Finally, we comment briefly on the curious drop in the electrical resistivity exponent from the marginal Fermi-liquid value at low pressures to a lower value of approximately 3/2 at higher pressures—that is, at pressures where the magnetic transition becomes first order and beyond (Fig. 4 inset). This surprising behaviour, which was reported in an earlier study by some of us²¹, is inconsistent with the predictions of the SCR model and remains a mystery that ZrZn_2 shares with other materials, including MnSi (refs 11, 24, 25) and PrNiO_3 (ref. 26). One possibility is that the state of these systems close to and beyond p_c (that is, in the first-order regime and above) is characterized by slow magnetic inhomogeneities not correctly described using the SCR model in its present form²⁷. Recent treatments that go beyond the SCR approximation indeed suggest that magnetic inhomogeneities might arise generically near a ferromagnetic quantum phase transition ($p \gtrsim p_c$)^{28–31}. Furthermore, energy-band structure calculations³² suggest, seemingly paradoxically, that antiferromagnetic correlations can be pronounced in nearly ferromagnetic metals similar to ZrZn_2 at $p \gtrsim p_c$, and can strongly modify the temperature dependence of the resistivity above p_c from that expected in a marginal Fermi-liquid description.

Received 11 July; accepted 29 August 2008.

- Landau, L. D. *Collected Papers* (ed. Ter Haas, D.) Ch. 90, 91 (Pergamon, 1965).
- Stewart, G. R. Non-Fermi-liquid behaviour in d and f-electron metals. *Mod. Phys.* **73**, 797–855 (2001).
- Baym, G. & Pethick, C. *Landau Fermi Liquid Theory* Ch. 3 (Wiley, 1991).
- Holstein, T., Norton, R. E. & Pincus, P. de Haas-van Alphen effect and the specific heat of an electron gas. *Phys. Rev. B* **8**, 2649–2656 (1973).
- Dzyaloshinskii, I. E. & Kondratenko, P. S. Theory of weak ferromagnetism in a Fermi fluid. *Sov. Phys. JETP* **43**, 1036–1054 (1976).
- Reizer, M. & Yu. Effective electron-electron interaction in metals and superconductors. *Phys. Rev. B* **39**, 1602–1608 (1989).
- Varma, C. M., Littlewood, P. B., Schmitt-Rink, S., Abrahams, E. & Ruckenstein, A. E. Phenomenology of the normal state of Cu-O high temperature superconductors. *Phys. Rev. Lett.* **63**, 1996–1999 (1989).
- Moriya, T. *Spin Fluctuations in Itinerant Electron Magnetism* (Springer, 1985).
- Lonzarich, G. G. in *Electron* (ed. Springford, M.) 109–147 (Cambridge Univ. Press, 1997).
- Millis, A. J. Effect of a nonzero temperature on quantum critical points in itinerant fermion systems. *Phys. Rev. B* **48**, 7183–7196 (1993).
- Pfleiderer, C., McMullan, G. J., Julian, S. R. & Lonzarich, G. G. Magnetic quantum phase transition in MnSi under hydrostatic pressure. *Phys. Rev. B* **55**, 8330–8338 (1997).
- Pfleiderer, C., Julian, S. R. & Lonzarich, G. G. Non-Fermi liquid nature of the normal state of itinerant-electron ferromagnets. *Nature* **414**, 427–430 (2001).
- Nicklas, M. *et al.* Non-Fermi-liquid behaviour at a ferromagnetic quantum critical point in $\text{Ni}_x\text{Pd}_{1-x}$. *Phys. Rev. Lett.* **82**, 4268–4271 (1999).
- Kuchler, R. *et al.* Thermal expansion and Grüneisen ratio near quantum critical points. *Physica B* **378–380**, 36–39 (2006).
- Uhlarz, M., Pfleiderer, C. & Hayden, S. M. Quantum phase transitions in the itinerant ferromagnet ZrZn_2 . *Phys. Rev. Lett.* **93**, 256404 (2004).
- Yates, S. J. C., Santi, G., Hayden, S. M., Meeson, P. J. & Dugdale, S. B. Heavy quasiparticles in the ferromagnetic superconductor ZrZn_2 . *Phys. Rev. Lett.* **90**, 057003 (2003).
- Lonzarich, G. G., Bernhoeft, N. R. & Paul, D. M. Spin density fluctuations in magnetic metals. *Physica B* **156–157**, 699–705 (1989).
- Bernhoeft, N. R., Law, S. A. & Lonzarich, G. G. Magnetic excitations in ZrZn_2 at low energies and long wavelengths. *Phys. Scripta* **38**, 191–193 (1988).
- Yelland, E. A. *et al.* Ferromagnetic properties of ZrZn_2 . *Phys. Rev. B* **72**, 184436 (2005).
- Sokolov, D. A., Aronson, M. C., Gannon, W. & Fisk, Z. Critical phenomena and the quantum critical point of ferromagnetic $\text{Zr}_{1-x}\text{Nb}_x\text{Zn}_2$. *Phys. Rev. Lett.* **96**, 116404 (2006).
- Takashima, S. *et al.* Robustness of non-Fermi-liquid behavior near the ferromagnetic critical point in clean ZrZn_2 . *J. Phys. Soc. Jpn* **76**, 043704 (2007).
- Niklowitz, P. G. *et al.* Spin-fluctuation-dominated electrical transport of Ni_3Al at high pressure. *Phys. Rev. B* **72**, 024424 (2005).
- Brando, M. *et al.* Logarithmic Fermi-liquid breakdown in NbFe_2 . *Phys. Rev. Lett.* **101**, 026401 (2008).
- Thessieu, C. *et al.* Field dependence of the magnetic quantum phase transition in MnSi. *J. Phys. Condens. Matter* **9**, 6677–6687 (1997).
- Doiron-Leyraud, N. *et al.* Fermi liquid breakdown of the paramagnetic phase of a pure metal. *Nature* **425**, 595–599 (2003).
- Zhou, J. S., Goodenough, J. B. & Dabrowski, B. Pressure induced non-Fermi liquid behaviour of PrNiO_3 . *Phys. Rev. Lett.* **94**, 226602 (2005).
- Pfleiderer, C. *et al.* Partial order in the non-Fermi-liquid phase of MnSi. *Nature* **427**, 227–231 (2004).
- Chubukov, A. V., Maslov, D. L. & Millis, A. J. Nonanalytic corrections to the specific heat of a three-dimensional Fermi liquid. *Phys. Rev. B* **73**, 045128 (2006).
- Chitov, C. Y. & Millis, A. J. First temperature corrections to the Fermi-liquid fixed point in two dimensions. *Phys. Rev. B* **64**, 054414 (2001).
- Belitz, D., Kirkpatrick, T. R. & Vojta, T. Nonanalytic behaviour of the spin susceptibility in clean Fermi systems. *Phys. Rev. B* **55**, 9452–9462 (1997).
- Gehring, G. A. Pressure induced quantum phase transitions. *Europhys. Lett.* (in the press); preprint at (<http://arxiv.org/abs/0711.2586>) (2007).
- Jeong, T., Kyker, A. & Pickett, W. E. Fermi velocity spectrum and incipient magnetism in TiBe_2 . *Phys. Rev. B* **73**, 115106 (2006).
- Ueda, K. & Moriya, T. Contribution of spin fluctuations to the electrical and thermal resistivities of weakly and nearly ferromagnetic metals. *J. Phys. Soc. Jpn* **39**, 605–615 (1975).

Acknowledgements We acknowledge financial support from the Royal Society, the EPSRC and St Catharine's College, Cambridge, and the Ministry of Education, Culture, Sports, Science and Technology of Japan (Grants-in-Aid for Scientific Research on Priority Areas and (S) from MEXT, Japan). We thank S. E. Rowley and I. R. Walker for experimental help and S. E. Rowley, C. M. Varma and P. B. Littlewood for discussions.

Author Contributions R.P.S. set up, performed and analysed the resistivity measurements at ambient pressure and under pressure, and carried out the model calculations. M.S. set up, performed and analysed the thermal conductivity measurements. S.S.S. assisted with measurements under pressure. N.K., S.T., M.N. and H.T. grew the ZrZn_2 crystals for the study. G.G.L. and N.K. designed the study. R.P.S. and G.G.L. wrote the paper. M.S. assisted in writing the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to R.P.S. (rps24@cam.ac.uk).

LETTERS

Low-speed fracture instabilities in a brittle crystal

J. R. Kermode¹, T. Albaret², D. Sherman³, N. Bernstein⁴, P. Gumbsch^{5,6}, M. C. Payne¹, G. Csányi⁷ & A. De Vita^{8,9}

When a brittle material is loaded to the limit of its strength, it fails by the nucleation and propagation of a crack¹. The conditions for crack propagation are created by stress concentration in the region of the crack tip and depend on macroscopic parameters such as the geometry and dimensions of the specimen². The way the crack propagates, however, is entirely determined by atomic-scale phenomena, because brittle crack tips are atomically sharp and propagate by breaking the variously oriented interatomic bonds, one at a time, at each point of the moving crack front^{1,3}. The physical interplay of multiple length scales makes brittle fracture a complex ‘multi-scale’ phenomenon. Several intermediate scales may arise in more complex situations, for example in the presence of micro-defects or grain boundaries. The occurrence of various instabilities in crack propagation at very high speeds is well known¹, and significant advances have been made recently in understanding their origin^{4,5}. Here we investigate low-speed propagation instabilities in silicon using quantum-mechanical hybrid, multi-scale modelling and single-crystal fracture experiments. Our simulations predict a crack-tip reconstruction that makes low-speed crack propagation unstable on the (111) cleavage plane, which is conventionally thought of as the most stable cleavage plane. We perform experiments in which this instability is observed at a range of low speeds, using an experimental technique designed for the investigation of fracture under low tensile loads. Further simulations explain why, conversely, at moderately high speeds crack propagation on the (110) cleavage plane becomes unstable and deflects onto (111) planes, as previously observed experimentally^{6,7}.

Until relatively recently, the process of brittle fracture has been primarily analysed using continuum mechanics techniques, which predict the driving force for crack propagation—the release of elastic energy per unit advancement as the crack progresses—to be a function of the applied load, the length of the crack and a geometry-dependent factor^{1,2,8,9}. The crack is assumed to propagate as long as its energy release rate, G , is equal to or larger than a material- and orientation-dependent critical value. Griffith⁸, who first carried out this analysis, equated the critical energy release rate below which crack propagation cannot take place with 2γ , the energy needed to create the new crack surfaces. This approach provides a good description of the fracture process for a straight crack in both static⁸ and dynamic regimes⁹, as shown by recent experiments on carefully prepared silicon single crystals^{7,10–13}.

More delicate issues such as the origin of crack path instabilities and materials’ resistance to fracture are, however, poorly understood. Even very simple atomistic models show that the discrete nature of the material will manifest itself in a higher resistance to fracture than predicted by Griffith’s criterion, owing to lattice (or bond) trapping^{4,14}. This effect depends on the detailed nature of the atomic interactions in the specific material^{4,15}, and may be very different

for different cleavage planes. For a given cleavage plane, it can lead to slow crack growth^{1,4,16} and anisotropy with respect to propagation along different crystallographic directions^{17,18} or even the sign of the propagation direction¹⁹.

Our simulations use the ‘learn-on-the-fly’ (LOTF) hybrid classical/quantum mechanical molecular dynamics method, which provides a nearly seamless coupling of the mobile quantum mechanical and the classical regions²⁰. We use the quantum mechanical chemical-bonding description, where such coupling is necessary near the crack tip. In the rest of the sample it is unnecessary and would be prohibitively expensive computationally. We first consider a (111) crack plane with the crack front aligned with the $[1\bar{1}0]$ direction, which is under a uniaxial $[111]$ tensile load (Fig. 1). Our simulations indicate that the K-field displacement predicted by continuum theory^{1,2} is never stable. Below a load of $G = 2.1 \text{ J m}^{-2}$, which is much lower than the Griffith critical load ($G_c = 2\gamma = 3.1 \text{ J m}^{-2}$, using a value for the surface energy computed with our quantum mechanical model), the crack closes. Above this load, a crack-tip reconstruction forms spontaneously by means of a bond rotation (Fig. 1c), resulting in the formation of five- and seven-member rings (see Supplementary Information). This relieves the stress along the main load axis in the immediate tip region, thus inducing additional lattice trapping. The reconstructed geometry of the crack tip is structurally similar to the Pandey reconstruction of the (111) surface²¹. As such, it could be viewed as the core of a dislocation that remains trapped at the crack tip, unable to induce large-scale plasticity at temperatures much lower than the brittle-to-ductile transition.

In simulations performed above the critical load for propagation, the slowly moving (111) $[1\bar{1}0]$ crack occasionally visits the reconstructed tip configuration as it propagates. Remarkably, for slow enough propagation, the $[111]$ oriented bonds in the six-member rings located one atomic layer below the initial crack plane and further along the propagation direction can break before the bridging bonds associated with the seven-member ring. Indeed, these are the bonds where the stress is mostly concentrated after the reconstruction. Every such reconstruction and bond-breaking event creates a downward $[111]$ step for the system oriented as in Fig. 1c, owing to the broken mirror symmetry of the crystal. This is at variance with the behaviour expected for propagation at high speeds (larger than those investigated here), where both upwards and downwards steps may occur²² and macroscopically smooth (111) cleavage surfaces are invariably observed²³. Our simulations thus predict a systematic deflection of the crack path from the (111) plane at sufficiently low velocities, whereas fast cracks are dynamically steered away from the instability. We speculate that the deflection process may start with a reconstruction event induced by the local slowing of a crack-front segment, perhaps due to an encounter with a crystal defect. However it starts, this process will initiate a positive-feedback ‘sinking’

¹Theory of Condensed Matter Group, Cavendish Laboratory, University of Cambridge, CB3 0HE, UK. ²Université de Lyon 1, LPMCN, CNRS, UMR 5586, F69622 Villeurbanne Cedex, France. ³Department of Materials Engineering, Technion - Israel Institute of Technology, Haifa 32000, Israel. ⁴Center for Computational Materials Science, Naval Research Laboratory, Washington, DC 20375-5343, USA. ⁵Institut für Zuverlässigkeit von Bauteilen und Systemen, Universität Karlsruhe (TH), Kaiserstrasse 12, 76131 Karlsruhe, Germany. ⁶Fraunhofer Institut für Werkstoffmechanik, Wöhlerstrasse 11, 79108 Freiburg, Germany. ⁷Engineering Laboratory, University of Cambridge, CB2 1PZ, UK. ⁸King’s College London, Department of Physics, Strand, London WC2R 2LS, UK. ⁹INFN-DEMOCRITOS National Simulation Center and Center of Excellence for Nanostructured Materials, University of Trieste, Trieste I-34127, Italy.

mechanism, because each reconstruction event will further trap and slow the crack, leading to ever more reconstructions, the whole process yielding observable fracture surface features.

A simple mesoscopic model of the crack-tip reconstruction and ensuing feedback process shows how this perturbation can evolve into the macroscopic ridge-like feature (Fig. 2d). A crack-front section of length L is modelled by an elastic string²⁴, represented by a set of N beads, each of mass m , connected by springs. The beads' positions and velocities are represented by the array of dynamical variables $\{\mathbf{r}_i, \mathbf{v}_i\}$, $i = 1, \dots, N$, and periodic boundary conditions are imposed in the crack-front y direction. The values of most of the parameters in what follows can be directly derived from experimental data, as detailed in the Supplementary Information. The equations of motion for the beads are expressed as the following dynamical map, where Δt is the time step associated with each iteration of the map:

$$\mathbf{r}_i(t + \Delta t) = \mathbf{r}_i(t) + \Delta t \mathbf{v}_i(t)$$

$$\mathbf{v}_i(t + \Delta t) = \mathbf{v}_i(t) + \frac{\Delta t}{m} [f^0 \hat{\mathbf{x}} - b \mathbf{v}_i(t) + k(\mathbf{r}_{i+1}(t) + \mathbf{r}_{i-1}(t) - 2\mathbf{r}_i(t)) - \xi_i \frac{\tau}{l} (v_i^x(t))^2 \hat{\mathbf{x}} + \eta \frac{L}{N} \sum_j \max(0, r_j^x - r_i^x) \hat{\mathbf{x}}] \quad (1)$$

$$\xi_i = \frac{s}{1 + \exp((v_i^x - v_0^x)/d)} \quad (2)$$

Consider the terms on the right-hand side of equation (1), in the order they appear. The model incorporates crack-front acceleration in the propagation direction x due to the driving force $f^0 \hat{\mathbf{x}}$ induced by

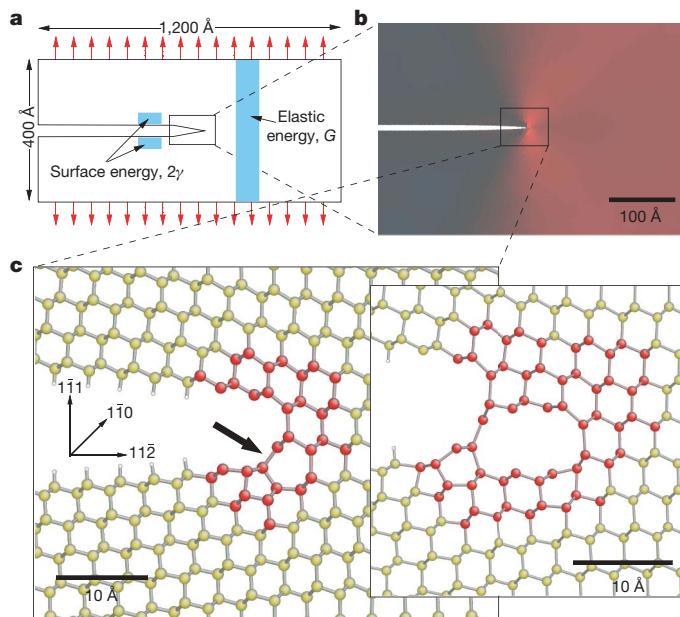


Figure 1 | The simulated (111)[110] crack system. **a**, Thin-strip geometry of the simulation. **b**, Map of the σ_{yy} stress component near the crack tip (red, highly stressed; black, unstressed). **c**, Ground-state reconstruction geometry of the crack tip under load, created by a rotation of the bond indicated by the arrow. Red atoms are described using quantum mechanics, yellow atoms using an interatomic potential. The small white atoms represent hydrogen atoms that we used to terminate the initial surface behind the crack to prevent it from closing up at low loads, and thereby ease the search for a load value that stabilizes the crack (with or without terminating hydrogen atoms). The first few bonds that break after the reconstruction are in six-member rings adjacent to the seven-member ring of the reconstruction. Inset, crack-tip structure after reconstruction and subsequent bond breaking. Further increasing the load finally breaks the crack-tip bridging bond and advances the open crack surface, yielding a downward step in the crack plane (not shown).

the load G . This continues until a limiting speed $v_{\max} = f^0/b$, due to a velocity-dependent phonon drag force $-b\mathbf{v}_i(t)$ with friction coefficient b , is reached. Line tension along the crack front is modelled by the springs of strength k . Crack-tip reconstruction occurs at a rate which smoothly rises for decreasing crack speeds, leading to the crack climbing by one atomic layer after each reconstruction and slowing the crack due to reconstruction-induced lattice trapping. This is the key assumption of the model and yields the drag term quadratic in the velocity, which is connected with equation (2) as discussed in detail below. Last, there is a driving force increase for crack-front sections 'held back' during the propagation, to include stress-enhancement effects at those segments of the crack that have slowed and trail behind the rest of the crack front. This leads to extra driving forces for these segments, which eventually level the ridges. Although the last effect is well known, its precise functional formulation and scaling are not. However, the model is not very sensitive to the details of the stress-enhancement term, so varying its functional form and prefactor η gives rise to only slightly different ridge shapes.

The variable ξ in equation (2) (dropping the subscript) is the local climb slope due to reconstruction, which is velocity dependent. The exact functional form of ξ is again not particularly important, as long as it implies no reconstruction at high velocities (where, thus, $\xi \approx 0$) and a smooth transition to the maximum observed climb slope, s , at low velocities. A simple exponential form with scale parameter d that centres the transition on a chosen value $v_0 < v_{\max}$ is used in equation (2), which thus just expresses the fact that climbing results from low propagation speeds, in agreement with our calculations and experiments. A macroscopic propagation instability can arise if low propagation speeds can, in turn, be caused by climbing, creating a positive-feedback loop, when a perturbation initiates the whole process by slowing a point of the crack front below a critical speed $v_c < v_{\max}$.

Our simulations provide an atomic-scale mechanism leading to such a situation, by showing that climbing is associated with tip reconstruction events, promoted by low propagation speeds and, in turn, capable of momentarily trapping, and thus locally further slowing, the crack front. In our model we represent this deceleration by imparting to the system a small impulse, directed opposite to the

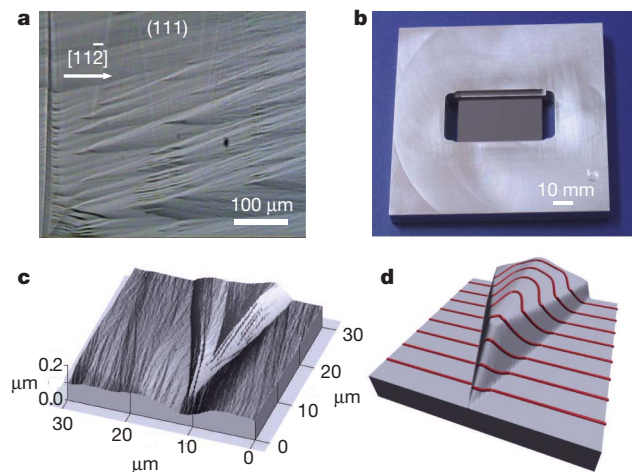


Figure 2 | Ridges formed by low-speed instabilities on the (111) crack plane. **a**, Optical micrograph of the (111) crack system. The crack advances from left to right as it accelerates up to a speed of about 800 m s^{-1} , as determined from the faintly visible Wallner lines³⁰. The ridges start to form at a variety of crack speeds. **b**, The fracture apparatus, showing the silicon specimen glued to the aluminium frame. **c**, Atomic force micrograph close-up of a ridge. **d**, Evolution of the crack front with the ridge shape predicted by a mesoscopic model based on the atomistic reconstruction mechanism illustrated in Fig. 1. The red lines represent snapshots of the propagating crack front yielding the interpolated crack surface (grey).

direction of motion, associated with each reconstruction event. If we assume for simplicity that an average of n reconstruction events take place during the time interval Δt for a given bead, its momentum mv will drop by $-nmv\tau/\Delta t$, where the parameter $\tau \ll \Delta t$ is representative of the time delay associated with each reconstruction. The number of events is related to the propagation velocity by $n = \xi v \Delta t / l$, where l is the bilayer distance in the (111) direction, that is, the amount of climb for each reconstruction event. Thus, the momentum change is $-m\xi v^2 \tau / l$, which is the nonlinear drag term in equation (1) discussed above. We note that if the frequency of reconstruction events were constant for a given temperature, as is the case for simple activated processes, then ξ would scale as $1/v$, and would just give rise to a simple drag term, linear in the velocity. Thus, in an indirect way, this model sheds light on the role of the cooperative, dynamic nature of brittle fracture: at high speeds, the atoms near the crack tip are steered towards clean cleavage.

We perform experimental studies of the low-crack-speed regime using a technique for applying very small but steady and well-controlled tensile loads. A silicon specimen is loaded by taking advantage of the thermal expansivity mismatch between the sample and the aluminium loading frame (Fig. 2b). Micrographs of the resulting (111) fracture surface are shown in Fig. 2a, c. Triangular ridges, all deviating in the same direction from the fracture surface, form at a range of low crack speeds below about 800 m s^{-1} . At higher crack speeds, of about $2,000 \text{ m s}^{-1}$, the surface is mirror smooth and no ridges are present. The crystallographic direction of the deviation (identical in over 40 independent samples) is the same as the reconstruction-induced steps in the atomistic simulation, and the shape of the ridges is qualitatively in agreement with the mesoscopic model (Fig. 2c, d). Similar features have recently been reported²³ under more complicated loading conditions at a speed of about $1,000 \text{ m s}^{-1}$ (see Supplementary Information).

We next considered the (110) crack plane. Experiments have shown that (110)[$\bar{1}\bar{1}0$] cracks propagating along the [001] direction deflect out of the plane at very low velocity^{6,18}, whereas (110)[001] cracks propagating in the [$\bar{1}\bar{1}0$] direction stay on the (110) plane up to very high velocity (about $2,900 \text{ m s}^{-1}$) before also faceting onto (111) planes^{6,7}. Recent studies²⁵ have assigned critical velocities for this instability for various propagation directions and have shown that in all cases the deflection is not immediate but only occurs after some initial propagation on the (110) plane. We simulated the (110)[$\bar{1}\bar{1}0$] crack propagation and observed the onset of this deflection. This crack propagates by breaking a series of bonds, labelled A in Fig. 3a. The system also contains type-B bonds, which are oriented at the same angle with respect to the tension axis as A bonds. Resolving the elastic stresses on the atomic scale reveals that B bonds located immediately above and below the crack plane are almost as highly stressed as A bonds. However, strong neighbour bonds connecting under-coordinated atoms left exposed by the advancing crack make the A bonds weaker than B bonds, so the former are expected to break selectively during quasi-static (110) cleavage, consistent with the low-speed experimental observations.

In the simulations we initially observe each newly exposed under-coordinated surface atom snap back towards the subsurface region. This induces significant local atomic motion and the excess energy diffuses into the bulk crystal, but no immediate rebonding occurs as long as the speed of propagation is sufficiently low. However, bond-breaking events become more frequent with increasing crack speed, and the local relaxations overlap in time. As more kinetic energy is locally available, the fast crack front 'stumbles'. Local reconstructions involving the removal of under-coordinated atoms begin to occur on the open surfaces. This removes the reason for selective A-bond breaking and in our simulation we indeed observe B-bond breaking events that deflect the crack front onto a (111) plane. Further simulations reveal that under these conditions any slight disturbance away from pure tension (for example a small extra shear strain component as in Fig. 3b) can systematically reverse the initial relative stability of

A and B bonds. This results in multiple coherent breaking of B bonds, effectively exposing (111) surfaces, consistent with the crack motion observed experimentally (Fig. 3c–e).

No near degeneracy of crack-tip bonds exists in the orthogonal [$\bar{1}\bar{1}0$] direction, consistent with the observation of instabilities only at much higher speeds⁶. The deflection mechanism is thus only

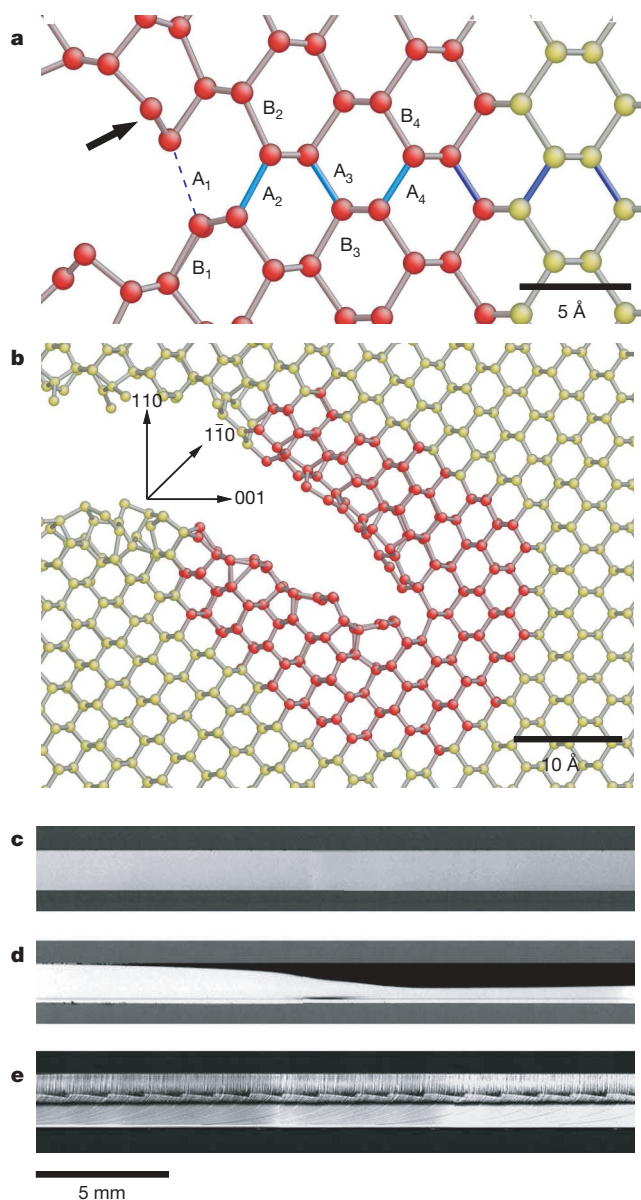


Figure 3 | The (110)[$\bar{1}\bar{1}0$] crack system. **a**, Geometry of the crack tip propagating straight at low speeds by sequential breaking of type A bonds (blue). Red atoms are described using quantum mechanics, yellow atoms using an interatomic potential. At each propagation step, the A bond is weaker than the corresponding B bond because of a neighbouring under-coordinated atom (indicated by an arrow for the A_1 – B_1 pair). **b**, At higher speed an instability occurs. At this point any slight shear disturbance in the stress field reverses the relative stability of A and B bonds, and the crack is deflected onto a (111) plane. The energy release rate for this system is $G = 6.7 \text{ J m}^{-2}$. In addition to the main tensile load, this includes a small $G = 0.24 \text{ J m}^{-2}$ shear contribution to break the symmetry in the y direction. Both exposed (111) crack surfaces undergo a 2×1 reconstruction. **c–e**, Photographs from three-point bending experiments⁶ (reprinted with permission): (110) fracture surface for low-speed cleavage (grey; **c**); fracture surface of a specimen with the crack deflected from the (110) plane to the (111) plane (black) for intermediate-speed cleavage (**d**); (111) fracture surface obtained for high-speed cleavage (**e**).

operative for the [001] crack propagation direction. By contrast with the (111) cleavage plane, here the formation of new chemical bonds, eventually leading to crack deflection, is favoured by a relatively high propagation speed rather than a low one. This is due to the different interplay between the local atomic motion and the 'lattice trapping energy' liberated by the cleavage during the time intervals between consecutive crack-tip advancement events.

For both cleavage planes in silicon, subtle atomistic details of the crack tip involving dynamical reconstruction mechanisms govern the path of the crack. This is unlikely to be the case for silicon only, owing to the general multi-scale character of the fracture processes. In the Supplementary Information we present data on preliminary simulations that indicate similar crack-tip reconstruction phenomena in diamond and silicon carbide, and on experiments on sapphire that show low-speed instabilities similar to those observed in silicon. Therefore, we suggest that crack-tip reconstructions occur more generally in brittle solids and control the qualitative macroscopic fracture behaviour.

METHODS SUMMARY

Tensile loading of the silicon (111) system was performed using apparatus shown in Fig. 2b. The load is applied by heating the whole set-up by a few degrees Celsius. The experimental setup for the (110) system was identical to that in ref. 6. The LOTF hybrid simulations were carried out using the Stillinger–Weber classical potential²⁶ and a quantum mechanical model based on density functional theory, as implemented in the CASTEP²⁷ and SIESTA²⁸ packages. The quantum mechanically treated region was selected automatically during the simulation, on the basis of geometric criteria in relation to the instantaneous crack-tip position.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 24 January; accepted 17 July 2008.

- Lawn, B. *Fracture of Brittle Solids* 2nd edn (Cambridge Univ. Press, 1993).
- Broberg, K. B. *Cracks and Fracture* Ch. 3, 4 (Academic, 1999).
- Lin, I. H. & Thomson, R. Cleavage, dislocation emission, and shielding for cracks under general loading. *Acta Metall.* **34**, 187–206 (1986).
- Marder, M. Effects of atoms on brittle fracture. *Int. J. Fract.* **130**, 517–555 (2004).
- Buehler, M. J. & Gao, H. J. Dynamical fracture instabilities due to local hyperelasticity at crack tips. *Nature* **439**, 307–310 (2006).
- Sherman, D. & Be'ery, I. From crack deflection to lattice vibrations—macro to atomistic examination of dynamic cleavage fracture. *J. Mech. Phys. Solids* **52**, 1743–1761 (2004).
- Cramer, T., Wanner, A. & Gumbsch, P. Energy dissipation and path instabilities in dynamic fracture of silicon single crystals. *Phys. Rev. Lett.* **85**, 788–791 (2000).
- Griffith, A. A. The phenomena of rupture and flow in solids. *Phil. Trans. R. Soc. Lond. A* **221**, 163–198 (1921).
- Freund, L. B. *Dynamic Fracture Mechanics* (Cambridge Univ. Press, 1990).
- Field, J. E. Brittle fracture: its study and application. *Contemp. Phys.* **12**, 1–31 (1971).
- Hauch, J. A., Holland, D., Marder, M. P. & Swinney, H. L. Dynamic fracture in single crystal silicon. *Phys. Rev. Lett.* **82**, 3823–3826 (1999).
- Be'ery, I., Lev, U. & Sherman, D. On the lower limiting velocity of a dynamic crack in brittle solids. *J. Appl. Phys.* **93**, 2429–2434 (2003).
- Gally, B. J. & Argon, A. S. Brittle-to-ductile transitions in the fracture of silicon single crystals by dynamic crack arrest. *Phil. Mag. A* **81**, 699–740 (2001).
- Thomson, R., Hsieh, C. & Rana, V. Lattice trapping of fracture cracks. *J. Appl. Phys.* **42**, 3154–3160 (1971).
- Sinclair, J. E. The influence of the interatomic force law and of kinks on the propagation of brittle cracks. *Phil. Mag.* **31**, 647–671 (1975).
- Cook, R. F. & Linegar, E. G. Kinetics of indentation cracking in glass. *J. Am. Ceram. Soc.* **76**, 1096–1105 (1993).
- Riedle, J., Gumbsch, P. & Fischmeister, H. F. Cleavage anisotropy in tungsten single crystals. *Phys. Rev. Lett.* **76**, 3594–3597 (1996).
- Perez, R. & Gumbsch, P. Directional anisotropy in the cleavage fracture of silicon. *Phys. Rev. Lett.* **84**, 5347–5350 (2000).
- van Eck, H., van der Laan, D., Dhallé, M., ten Haken, B. & ten Kate, H. Critical current versus strain research at the University of Twente. *Supercond. Sci. Technol.* **16**, 1026–1030 (2003).
- Csányi, G., Albaret, T., Payne, M. & De Vita, A. "Learn on the fly": A hybrid classical and quantum-mechanical molecular dynamics simulation. *Phys. Rev. Lett.* **93**, 175503 (2004).
- Pandey, K. C. New π -bonded chain model for Si(111)–(2×1) surface. *Phys. Rev. Lett.* **47**, 1913–1917 (1981).
- Hoshi, T., Igushi, T. & Fujiwara, T. Nanoscale structures formed in silicon cleavage studied with large-scale electronic structure calculations: Surface reconstruction, steps, and bending. *Phys. Rev. B* **72**, 075323 (2005).
- Sherman, D., Markovitz, M. & Barkai, O. Dynamic instabilities in 111 silicon. *J. Mech. Phys. Solids* **56**, 376–387 (2008).
- Lange, F. F. The interaction of a crack front with a second-phase dispersion. *Phil. Mag.* **22**, 983–992 (1970).
- Sherman, D. Macroscopic and microscopic examination of the relationship between crack velocity and path and Rayleigh surface wave speed in single crystal silicon. *J. Mech. Phys. Solids* **53**, 2742–2757 (2005).
- Stillinger, F. H. & Weber, T. A. Computer simulation of local order in condensed phases of silicon. *Phys. Rev. B* **31**, 5262–5271 (1985).
- Segall, M. D. et al. First-principles simulation: ideas, illustrations and the CASTEP code. *J. Phys. Condens. Mater.* **14**, 2717–2743 (2002).
- Sanchez-Portal, D., Ordejón, P., Artacho, E. & Soler, J. M. Density-functional method for very large systems with LCAO basis sets. *Int. J. Quantum Chem.* **65**, 453–461 (1997).
- Kwon, I., Biswas, R., Wang, C., Ho, K. & Soukoulis, C. Transferable tight-binding models for silicon. *Phys. Rev. B* **49**, 7242–7250 (1994).
- Sharon, E., Cohen, G. & Fineberg, J. Propagating solitary waves in a rapidly moving crack front. *Nature* **410**, 68–71 (2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements P.G. acknowledges support from the Deutsche Forschungsgemeinschaft (Gu 367/30). N.B. acknowledges support from NRL and ONR. D.S. acknowledges support from the ISF (grant no. 1110/04). J.R.K., G.C. and M.C.P. acknowledge support from the EPSRC portfolio grant GR/S61263/01. T.A. acknowledges support from ANR–France (grant ANR-05-CIGC:LN3M) and IDRIS (Orsay, France, project 051841). A.D.V. acknowledges support from the EPSRC grant EP/SC23938/1. G.C. acknowledges support from the EPSRC grant EP/C52392X/1. The authors thank A. Sutton for a critical reading of the manuscript. Computer time was in part provided by the US Department of Defense HPCMP and the HPCS at the University of Cambridge.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to G.C. (gc121@cam.ac.uk).

METHODS

For the (111) system, cleaving of 525- μm -thick silicon specimens of $42 \times 26\text{-mm}^2$ lateral dimensions, cut from a [110] 4-inch silicon wafer, was performed under tensile load using the apparatus shown in Fig. 2b. The specimens were notched with a 150- μm -thick diamond saw to a length of 1.5 mm, heated to 200 °C and quenched in water to generate a pre-crack of the required length. The specimens were then glued to an A17075 T6 frame using a 150- μm -thick film of Epon 815C epoxy resin. The load was applied by heating the whole set-up by a few degrees Celsius on a laboratory heating stage, using the fact that the thermal expansion coefficients of the sample and the loading frame are different. When the rate of heating was well below 1 °C min⁻¹, the instabilities were always present. Using high rates of heating resulted in the propagation of fast cracks leaving behind mirror-smooth surfaces.

In the computer simulations, the CASTEP²⁷ package was used for the (111) system and the SIESTA²⁸ package for the (110) system. The classical potential was adjusted to match the linear elastic properties of the quantum mechanical models in each case. The simulations were carried out in the plane strain configuration and free boundary conditions were used in the x direction. The load, applied by imposing a fixed-displacement boundary condition in the y direction, was increased in a number of steps starting from $G = 2.5 \text{ J m}^{-2}$, allowing at least 1-ps equilibration time between successive steps, until the crack started to propagate. A variety of strain rates were tested down to 0.1% ps⁻¹, with no qualitative difference observed. The load needed to crack the system was 6.0 J m⁻² at the lowest strain rate we could achieve. Because the load was found still to be strongly strain rate dependent even at this strain rate, this value is only an upper bound.

The constant temperature (300 K) simulations for the system shown in Fig. 1 were made on a $1200 \times 400 \times 3.8\text{-}\text{\AA}^3$, 90,868-atom model thin slab, periodic in the crack-front direction. For these low-temperature, highly driven systems, thermal activation and entropy effects, which could induce a qualitatively different behaviour in three-dimensional systems, are unlikely to have a significant role. A series of quasi-static simulations show the same behaviour, although the system has to be overloaded much more in order to crack, as expected owing to the high lattice trapping.

In the plane-wave density functional theory part of the LOTF hybrid simulation, the Brillouin zone was sampled with four points along the crack front, exchange and correlation were described using the Perdew–Burke–Ernzerhof functional, nuclei were described by ultrasoft pseudopotentials, and the electronic wavefunctions were expanded in plane waves with a 180-eV kinetic energy cut-off. The quantum mechanical region was selected using an algorithm detailed in the Supplementary Information. The buffer region of the LOTF scheme²⁰ was set at about 8 \AA (fourth-neighbour distance).

Furthermore, we made a test calculation on a 15.2- \AA -thick slab using a tight binding²⁹ quantum mechanical engine to verify that the same phenomena apply in the true three-dimensional case. The reconstruction occurred at all sites along the crack front in quick succession before any other kind of rebonding took place. The simulation of the system shown in Fig. 3 was made on a $687 \times 655 \times 7.6\text{-}\text{\AA}^3$, 174,752-atom model slab, using the standard single zeta plus polarization basis set with soft confinement for the local orbital density functional theory part of the calculation. The mesh cut-off was 50 Ry (1 Ry = 13.6 eV) and norm-conserving pseudopotentials were used.

LETTERS

Medieval forewarning of the 2004 Indian Ocean tsunami in Thailand

Kruawun Jankaew¹, Brian F. Atwater², Yuki Sawai³, Montri Choowong¹, Thasinee Charoentitirat¹, Maria E. Martin⁴ & Amy Prendergast⁵

Recent centuries provide no precedent for the 2004 Indian Ocean tsunami, either on the coasts it devastated or within its source area. The tsunami claimed nearly all of its victims on shores that had gone 200 years or more without a tsunami disaster¹. The associated earthquake of magnitude 9.2 defied a Sumatra–Andaman catalogue that contains no nineteenth-century or twentieth-century earthquake larger than magnitude 7.9 (ref. 2). The tsunami and the earthquake together resulted from a fault rupture 1,500 km long that expended centuries' worth of plate convergence^{2–5}. Here, using sedimentary evidence for tsunamis⁶, we identify probable precedents for the 2004 tsunami at a grassy beach-ridge plain 125 km north of Phuket. The 2004 tsunami, running 2 km across this plain, coated the ridges and intervening swales with a sheet of sand commonly 5–20 cm thick. The peaty soils of two marshy swales preserve the remains of several earlier sand sheets less than 2,800 years old. If responsible for the youngest of these pre-2004 sand sheets, the most recent full-size predecessor to the 2004 tsunami occurred about 550–700 years ago.

The 2004 Indian Ocean tsunami, cresting higher in Thailand than it did anywhere else east of Sumatra (Fig. 1b), rose as much as 20 m above sea level on Phra Thong Island⁷. The main wave, as observed from a hill on the island's western shore (H, Fig. 1c), formed a relentless flood that rose stepwise to heights above treetops⁸. The tsunami ran more than 2 km inland across a Holocene plain composed of grassy beach ridges and intervening tree-lined swales.

The flooding at Phra Thong, as on other Thai coastal plains^{9,10}, produced local erosion and widespread deposition. The tsunami reamed out several drainages previously cut across beach ridges as much as 300 m inland (Fig. 1c, e, and Supplementary Fig. 1). In addition, it coated most of the island's western half with a sheet of sand (Fig. 1d). This sheet is locally lineated with sand streaks that extend inland from the spoil piles of tin miners (Fig. 1e and Supplementary Fig. 1). Its mean particle size ranges from medium to very fine on a transect across the northern part of the island¹¹. Horizontal bedding is common, as is overall upward fining to coarse silt.

Building on a previous reconnaissance¹², we sought pre-2004 sand sheets at Phra Thong by digging pits and augering holes into ridges and swales at more than 150 sites (Fig. 1d (dots) and Supplementary Table 1). At 20 of these sites we found pre-2004 sand interbedded with the peaty soils of swales that hold standing water most of the year (Fig. 1d, red dots). We found no pre-2004 sand beds in the quartz-sand soils of the ridges or in the slightly organic soils of swales that are merely damp.

We traced pre-2004 beds across each of two marshy swales near a place where the 2004 tsunami reportedly flowed about 10 m deep

(Figs 1e and 2a, and Supplementary Fig. 3a). These swales formed about 2,500 years ago (bark and shell dates; Fig. 2b, c), when the area's relative sea level was probably within 1–2 m of its present position¹³. Because beaches have built the island westwards, the more westerly of the swales (X) postdates its neighbour (Y). We assembled stratigraphic cross-sections from correlated pits, from auger borings and from a trench 35 m long, estimated particle size in the field, inferred a preliminary chronology from radiocarbon dating of individual plant remains and shells (Figs 2c and 3, and Supplementary Table 2), and made diatom analyses (Supplementary Figs 4 and 5).

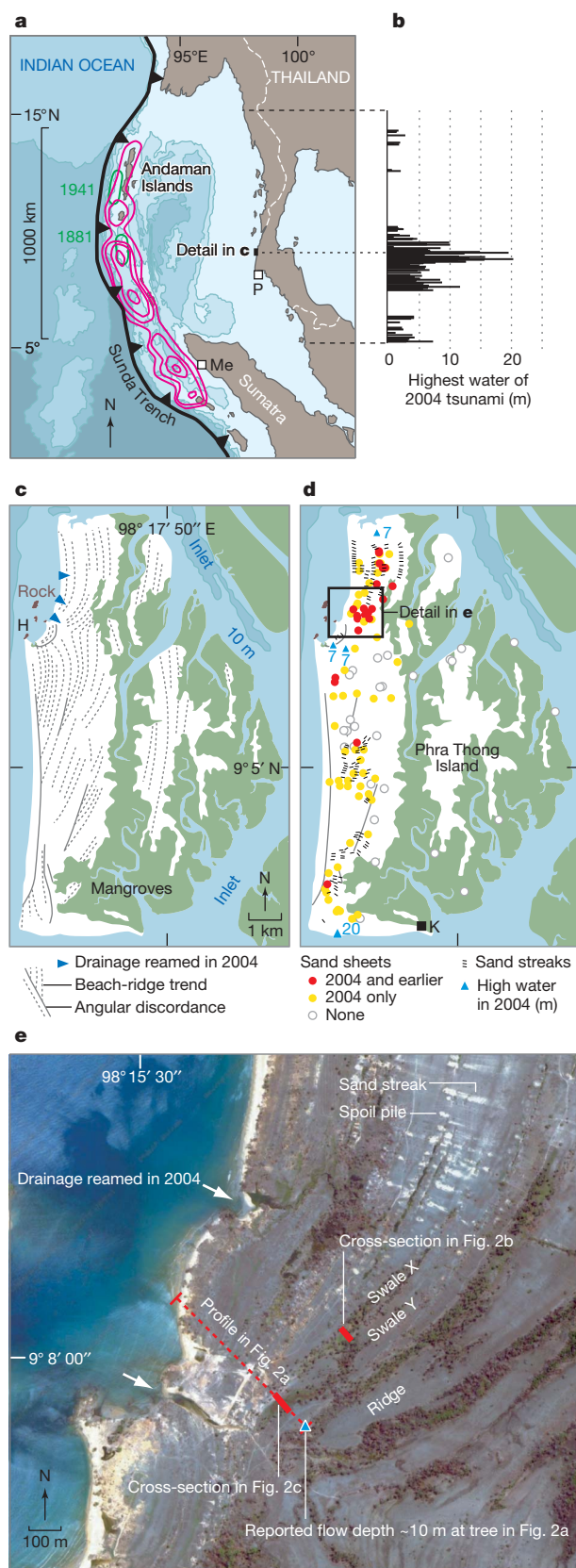
Peaty soil in swale X contains two sand sheets (B and C in Fig. 2b, d) that resemble the overlying 2004 deposit. Sheet C, the earlier, is commonly 10 cm thick. Coarse to very coarse sand forms a discontinuous basal layer that fills pre-existing pockets in the underlying soil. The rest of sheet C consists of very fine sand and coarse silt that contains horizontal laminae defined most visibly by leaf fragments (Fig. 3c). The entire sheet formed after 2,200–2,400 sidereal years ago, the age of an isolated piece of bark in the uppermost 1 cm of the underlying buried soil. Leaf fragments within the sheet gave ages that conflict with one another and that exceed the bark age by thousands of years (Fig. 3c). Sheet B, commonly 5 cm thick, typically fines upwards from fine sand to sandy silt. It conformably overlies peaty soil that contains a horizon of bark fragments in its uppermost 1 cm. Three of these fragments, collected separately, yielded ages between 530 ± 40 and 570 ± 40 radiocarbon years before AD 1950 (^{14}C yr BP; Fig. 3b). If scarcely younger than these fragments, sheet B was deposited about 550–700 sidereal years ago (Supplementary Table 2).

Three pre-2004 sand sheets alternate with peaty soil in swale Y (Fig. 2c, e). All three are similar in thickness to the overlying 2004 sand sheet, and all extend preferentially up the swale's seaward side. All were formed after the swale ceased holding an intertidal flat that is marked by non-abraded molluscan shells 2,500–2,800 sidereal years old (Fig. 2c and Supplementary Table 2). The lowest two sheets, otherwise undated and thus left uncorrelated with swale X, consist mainly of very fine to fine sand. They lack sedimentary structures, probably because of bioturbation that blurs their contacts with the soils beneath (Fig. 2e). The highest pre-2004 sheet (B) retains a sharp base and tabular shape that extend the full length of the trench (Fig. 2f). This sheet typically fines upwards from basal fine or medium sand to parallel-laminated very fine sand that abounds in leaf fragments (Fig. 3a). It probably correlates with sheet B of swale X because each is the youngest pre-2004 sand sheet in its swale and because the leaf fragments in swale Y yielded ages too young for correlation with sheet C (Fig. 3a, c).

Although the 2004 sand sheet abounds in brackish and marine diatoms, the earlier sand sheets in swales X and Y lack diatoms of

¹Department of Geology, Faculty of Science, Chulalongkorn University, Phayathai Road, Phatumwan, Bangkok 10330, Thailand. ²US Geological Survey at Department of Earth and Space Sciences, University of Washington, Seattle, Washington 98195-1310, USA. ³Geological Survey of Japan, National Institute of Advanced Industrial Science and Technology, Site C7 1-1-1 Higashi, Tsukuba 305-8567, Japan. ⁴Department of Earth and Space Sciences, University of Washington, Box 351310, Seattle, Washington 98195-1310, USA. ⁵Geoscience Australia, GPO Box 378, Canberra, Australian Capital Territory 2601, Australia.

any kind (Supplementary Figs 4 and 5). Marine and brackish-water diatoms aid in identifying tsunami deposits on temperate shores¹⁴. Perhaps their opaline silica valves do not last long in tropical warmth; in experiments, the dissolution of diatoms increases with temperature¹⁵.



Preservation is also a problem for the sand sheets themselves. The pre-2004 sheets are distinct and sharply bounded where the swale soil is peaty, blurred by gradational contacts where the soil is just slightly organic, and totally absent in the sandy soils of beach-ridge crests. The 2004 tsunami deposit is already headed towards this fate: in wet swales it has a protective cap of organic matter as much as 5 cm thick, whereas on ridge crests it lacks any cover other than ejecta from burrows that tap the underlying sandy soil. Sheet B, if truly correlative between swales X and Y, initially spanned the intervening beach ridge for a total shore-normal length of no less than 100 m (Fig. 1e).

Although sand sheets can record intense storms that drive waves over or through sandy beach berms¹⁶, the geographic setting limits Phra Thong's exposure to such storms. Less than 10° from the Equator, this part of Thailand fringes the belt where the Coriolis minimum limits cyclonic winds¹⁷. Scores of twentieth-century cyclones originated in Indian Ocean waters to its west, but all these moved towards India, Bangladesh or Myanmar¹⁸ without producing a known storm surge in Thailand¹⁹. Tropical cyclones do strike Thailand from its Pacific side. However, such a storm loses strength during its overland crossing to the Indian Ocean (an example is given in Supplementary Fig. 6), and its anticlockwise winds can pile the sea against Thailand's west coast only in the storm's trailing-left quadrant.

Phra Thong's setting also disfavours sand-sheet deposition by river or wind. Tidal inlets separate the island from the nearest rivers (Fig. 1c). Aeolian dunes obscure little, if any, of the island's delicate striping by beach ridges and swales (Fig. 1e and Supplementary Fig. 1).

Chronology provides three further reasons to ascribe the pre-2004 sand sheets to tsunamis. First, the middle Holocene ages of the leaf fragments from sheet C (Fig. 3) imply scour into long-buried deposits beneath tidal inlets. The 2004 tsunami showed capacity for such scour by knocking down mangroves along an inner part of the inlet that bounds Phra Thong Island on the south (at K in Fig. 1d). Second, the sand sheets represent infrequent events: the soil between sheets C and B spans 1,500–1,850 years, although it may contain the bioturbated remains of an intervening sheet (Supplementary Fig. 2c); and the interval between sheet B and the 2004 tsunami lasted nearly 550–700 years (ranges computed from calibrated ages in Supplementary Table 2). These time intervals are in the broad range of deductive estimates for the recurrence of giant earthquakes in the Sumatra-Andaman source region of the 2004 tsunami^{2–5}. Third, sheet B, if little younger than AD 1300–1450 (Supplementary Table 2), may correlate with tsunami and earthquake evidence elsewhere. The youngest widespread pre-2004 sand sheet on a beach-ridge plain at Meulaboh, Sumatra (Fig. 1a, Me) overlies plant detritus dated to AD 1290–1400 (ref. 20). Two coral fragments on a marine terrace in the Andaman Islands gave ages in the range AD 1200–1650 (ref. 21). However, in accounts from Ibn Battuta (journey, AD 1325–1354)²² and the great Ming armadas (voyages, AD 1405–1433)²³, we found no written evidence for a sheet-B tsunami on Sumatran and Sri Lankan shores that the 2004 tsunami would overrun.

Figure 1 | Setting. **a**, Northern Sunda Trench and vicinity. Red lines show a modelled fault slip during the 2004 Sumatra-Andaman earthquake⁵ at 5 (outer contour), 10, 15 and 20 m. Green lines show pre-2004 rupture areas^{2,26}. Bathymetry is shaded at 1-km intervals. Me, Meulaboh; P, Phuket. **b**, Heights of the 2004 tsunami on the eastern rim of the Indian Ocean^{7,27–30}. The tallest bars for Thailand obscure dozens of height measurements below local maxima. **c**, Landforms of Phra Thong Island. White areas denote beach ridge plains (examples in **e** and Supplementary Fig. 1). The lines of angular discordance (repeated for reference in **d**) probably record shoreline retreat. H, Hornbill Hill⁸. **d**, Tsunami deposits on Phra Thong Island. K, knocked-down mangroves along southern inlet. The 2004 tsunami heights are maxima in metres above sea level (ref. 7 and Supplementary Fig. 1). **e**, Area of stratigraphic evidence shown in Figs 2 and 3. Post-tsunami image, probably taken early in 2005, from PointAsia.com. Trees and shrubs (dark green) delineate some of the swales.

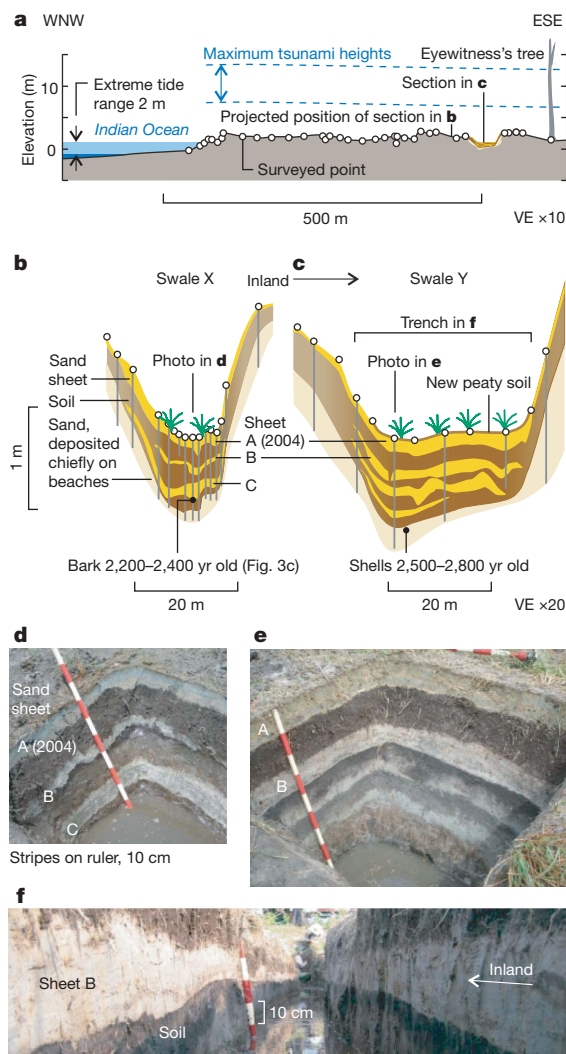


Figure 2 | Cross-sectional shape of sand sheets. **a**, Topographic profile along the line in Fig. 1e. VE, vertical exaggeration. Maximum tsunami heights are from an eyewitness in the indicated tree (Fig. 1e and Supplementary Fig. 3a) and from a post-tsunami survey 1 km to the southwest (Fig. 1d and Supplementary Fig. 1b). **b**, **c**, Cross-sections in swales X (**b**) and Y (**c**) from inferred correlation between pits and auger borings (vertical grey lines). Soil peaty in swales (dark brown) and sandy on ridges (light brown). Ages from Supplementary Table 2. **d**, **e**, Sand sheets alternate with dark peaty soils on the walls of pits in swales X (**d**) and Y (**e**). **f**, Lateral continuity of sand sheet B exposed in trench.

What tsunami sources might Phra Thong's pre-2004 sand sheets represent? Too little is known about the sheets' landward extent on the island, let alone their potential correlates on other Indian Ocean shores, to require full-size predecessors to the 2004 Sumatra–Andaman earthquake. However, the sheets probably required ruptures larger than that of 1881 (Fig. 1a); no sand sheet from the 1881 tsunami, which crested less than 1 m high on Indian tide gauges²⁴, is evident at Phra Thong Island in fibrous peaty soils that the 2004 tsunami failed to incise while covering them with sand (Fig. 2d and Supplementary Fig. 3c–g). The pre-2004 sheets may also require Sunda Trench earthquakes larger than magnitude 8.5 if, as estimated from numerical simulations³, such earthquakes would spawn Thai tsunamis only a few metres high—barely high enough to invade Phra Thong's beach-ridge plain (Fig. 2a).

Sand sheets of Phra Thong Island thus forewarn of infrequent catastrophe. They are already providing public officials and coastal residents with tangible evidence that the 2004 tsunami was not the first of its kind (Supplementary Fig. 7). Still to be determined is

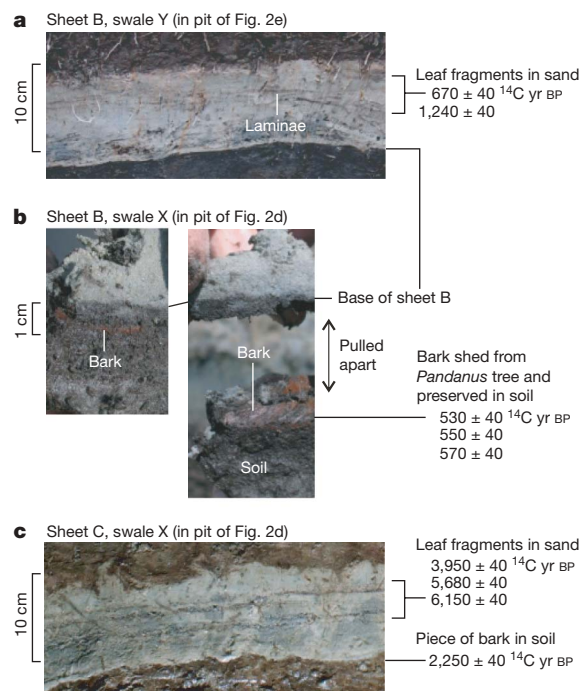


Figure 3 | Parallel laminae and radiocarbon ages. **a**, **b**, Indistinct horizontal laminae in sheet B are highlighted by leaf fragments. The two fragments dated yielded ages that differ from one another by about 600 radiocarbon years (**a**). The younger of these ages overlaps with ages on detrital tree-trunk bark in the highest part of the underlying soil (**b**). **c**, Leaf-fragments from laminae in sheet C yielded widely scattered ages thousands of years greater than that of bark in the underlying soil, all sampled within the same pit (Fig. 2d).

whether centuries dependably separate such outsize tsunamis of Sumatra–Andaman source, and whether these recur often enough to dominate Thailand's probabilistic tsunami hazard. Tsunamis without precedent in written history may threaten Indian Ocean shores that face other parts of the Sunda Trench and the Makran subduction zone^{24,25}. It can be hoped that natural warnings from recent geological history will help avert surprises from these additional tsunami sources.

METHODS SUMMARY

Landforms in Fig. 1c were traced from 1:50,000-scale aerial photographs taken in 1999 and from post-tsunami satellite images at PointAsia.com and Google Earth. In Fig. 2a the levelling had a closure error of 10 cm, and the tide levels refer to the Khura Buri gauge, 16 km northeast of Phra Thong. Diatom separation and analyses are described in Supplementary Figs 4 and 5. A separate suite of samples from sheet B yielded rare foraminiferal tests, all well preserved, that we do not interpret because of possible laboratory contamination with tests from the 2004 deposit. The radiocarbon age ranges reported in sidereal years span the 95% confidence interval from counting and calibration statistics (Supplementary Table 2).

Received 6 March; accepted 27 August 2008.

- Dominey-Howes, D., Cummins, P. & Burbidge, D. Historic records of teletsunamis in the Indian Ocean and insights from numerical modeling. *Nat. Hazards* **42**, 1–17 (2007).
- Bilham, R., Engdahl, R., Feldl, N. & Satyabala, S. P. Partial and complete rupture of the Indo-Andaman Plate boundary 1847–2004. *Seismol. Res. Lett.* **76**, 299–311 (2005).
- Lovholt, F. et al. Earthquake related tsunami hazard along the western coast of Thailand. *Nat. Hazards Earth Syst. Sci.* **6**, 979–997 (2006).
- Stein, S. & Okal, E. A. Ultralong period seismic study of the December 2004 Indian Ocean earthquake and implications for regional tectonics and the subduction process. *Bull. Seismol. Soc. Am.* **97**, S279–S295 (2007).
- Chlieh, M. et al. Coseismic slip and afterslip of the great M_w 9.15 Sumatra–Andaman earthquake of 2004. *Bull. Seismol. Soc. Am.* **97**, S152–S173 (2007).
- Bourgeois, J. in *The Sea* Vol. 15 (eds Bernard, E. N. & Robinson, A. R.) (Harvard Univ. Press, in the press).

7. Tsuji, Y. *et al.* The 2004 Indian tsunami in Thailand; surveyed runup heights and tide gauge records. *Earth Planets Space* **58**, 223–232 (2006).
8. Lyall, K. *Out of the Blue: Facing the Tsunami* (Australian Broadcasting Corporation, 2006).
9. Hori, K. *et al.* Horizontal and vertical variation of 2004 Indian tsunami deposits; an example of two transects along the western coast of Thailand. *Mar. Geol.* **239**, 163–172 (2007).
10. Choowong, M. *et al.* 2004 Indian Ocean tsunami inflow and outflow at Phuket, Thailand. *Mar. Geol.* **248**, 179–192 (2008).
11. Fujino, S. *et al.* in *Tsunamiites—Features and Implications* (eds Shiki, T., Tsuji, Y., Minoura, K. & Yamazaki, T.) 123–132 (Elsevier, 2008).
12. Fujino, S. *et al.* in *Proc. Int. Symp. on Restoration Program from Giant Earthquakes and Tsunamis* (ed. Kato, T.) 115–121 (Earthquake Research Institute, University of Tokyo, 2008).
13. Horton, B. P. *et al.* Holocene sea levels and palaeoenvironments, Malay-Thai Peninsula, Southeast Asia. *Holocene* **15**, 1199–1213 (2005).
14. Hemphill-Haley, E. Diatoms as an aid in identifying late-Holocene tsunami deposits. *Holocene* **6**, 439–448 (1996).
15. Kamatani, A. Dissolution rates of silica from diatoms decomposing at various temperatures. *Mar. Biol.* **68**, 91–96 (1982).
16. Donnelly, J. P., Butler, J., Roll, S., Wengren, M. & Webb, T. III. A backbarrier overwash record of intense storms from Brigantine, New Jersey. *Mar. Geol.* **210**, 107–121 (2004).
17. Gray, W. M. in *Storms* Vol. 1 (eds Pilke, R. J. & Pilke, R. S.) 145–163 (Routledge, 2000).
18. Pant, G. B. & Rupa Kumar, K. *Climates of South Asia* 320 (John Wiley & Sons, Chichester, England, 1997).
19. Murty, T. S. & Flather, R. A. Impact of storm surges in the Bay of Bengal. *J. Coast. Res.* **12** (Spec. Iss.), 149–161 (1994).
20. Monecke, K. *et al.* A 1,000-year sediment record of tsunami recurrence in northern Sumatra. *Nature* doi:10.1038/nature07374 (this issue).
21. Rajendran, K. *et al.* Age estimates of coastal terraces in the Andaman and Nicobar Islands and their tectonic implications. *Tectonophysics* **455**, 53–60 (2008).
22. Ibn Battuta. *The Travels of Ibn Battūta, A.D. 1325–1354* vol. 4 (transl. Gibb, H. A. R.) (Hakluyt Society, 1994).
23. Dreyer, E. L. *Zheng He: China and the oceans in the early Ming dynasty 1405–1433* (Pearson Longman, 2007).
24. Okal, E. A. & Synolakis, C. E. Far-field tsunami hazard from mega-thrust earthquakes in the Indian Ocean. *Geophys. J. Int.* **172**, 995–1015 (2008).
25. Heidarzadeh, M. *et al.* Historical tsunami in the Makran Subduction Zone off the southern coasts of Iran and Pakistan and results of numerical modeling. *Ocean Eng.* **35**, 774–786 (2008).
26. Ortiz, M. & Bilham, R. Source area and rupture parameters of the 31 December 1881 $M_w = 7.9$ Car Nicobar earthquake estimated from tsunamis recorded in the Bay of Bengal. *J. Geophys. Res.* **108** (B4) 2215, doi:10.1029/2002JB001941 (2003).
27. Satake, K. *et al.* Tsunami heights and damage along the Myanmar coast from the December 2004 Sumatra–Andaman earthquake. *Earth Planets Space* **58**, 243–252 (2006).
28. Siripong, A. Andaman seacoast of Thailand field survey after the December 2004 Indian Ocean tsunami. *Earthq. Spectra* **22**, 187–202 (2006).
29. Choi, B. H. Analysis and modeling of the distribution functions of runup heights of the December 26, 2004 earthquake tsunami in the Indian Ocean. (http://wave.skku.ac.kr/tsunami_survey_data/KEERC.report.pdf).
30. Hawkes, A. D. *et al.* Sediments deposited by the 2004 Indian Ocean tsunami along the Malaysia–Thailand Peninsula; Quaternary land–ocean interactions; sea-level change, sediments and tsunami. *Mar. Geol.* **242**, 169–190 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank B. Korsakun for logistical help; C. Tongjeen for permission for digging; V. Chutakositkanon, V. Jittanoon, T. Machado, T. Napradit, S. Pailoplee, S. Phantuwongraj, N. Rajeshwara Rao, S. Srinivasalu, P. Surakiatchai and A. Weerahong for contributions to Phra Thong field and laboratory work; Y. Fujii for providing bathymetric data; and S. Bondevik, M. Cisternas, H. Kelsey, A. Meltzner, K. Sieh, M. Tuttle and J. Woodruff for reviews. This report evolved from surveys supported by the Ministry of Natural Resources and Environment (Thailand), the National Science Foundation (USA), and the US Agency for International Development (participants are listed in Supplementary Table 3). Additional funding was provided by the Japan Society for the Promotion of Science (to Y.S.), the Thailand Research Fund (to M.C.), and Chulalongkorn University (through P. Charusiri).

Author Contributions All authors participated in the fieldwork, led by K.J. (Supplementary Table 1). B.A. prepared most of the manuscript. Y.S. analysed the diatoms and prepared several of the figures.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to K.J. (kjankaew@yahoo.co.uk).

LETTERS

A 1,000-year sediment record of tsunami recurrence in northern Sumatra

Katrin Monecke^{1†}, Willi Finger², David Klarer³, Widjo Kongko^{4†}, Brian G. McAdoo⁵, Andrew L. Moore⁶ & Sam U. Sudrajat⁷

The Indian Ocean tsunami of 26 December 2004 reached maximum wave heights of 35 m in Aceh, the northernmost province of Sumatra^{1,2}. Both the tsunami and the associated Sumatra–Andaman earthquake were unprecedented in Acehnese history^{3,4}. Here we use sand sheets to extend tsunami history 1,000 years into Aceh's past. The 2004 tsunami deposited a sand sheet up to 1.8 km inland on a marshy beach ridge plain. Sediment cores from these coastal marshes revealed two older extensive sand sheets with similar sediment characteristics. These sheets, deposited soon after AD 1290–1400 and AD 780–990, probably resulted from earlier tsunamis. An additional sand sheet of limited extent might correlate with a documented smaller tsunami of AD 1907. These findings, a first step towards a palaeotsunami record for northern Sumatra, suggest that damage-causing tsunamis in Aceh recur infrequently enough for entire human lifetimes to typically elapse between them. Such recurrence adds to the challenge of preparing communities along the northern Indian Ocean shorelines for future tsunamis.

Aside from the 2004 tsunami and its successor in 2005, few tsunamis are documented in 400 years of Indonesian history to have reached or approached Aceh (Fig. 1a)^{3–6}. The largest of these, in 1907, devastated the west coast of Simeulue Island and extended more than 950 km along the northwestern coast of Sumatra⁴. Run-up heights in Aceh are uncertain but, as estimated from interviews with inhabitants, they were far from reaching the extreme wave heights of the 2004 tsunami (Supplementary Note 1). Previously, in 1861, Aceh apparently sustained little or no damage from a documented west Sumatran tsunami⁴ associated with a likely predecessor of the Nias earthquake (moment magnitude (M_w) = 8.7) in March 2005 (ref. 7). The latter caused a tsunami that reached heights of 1–4 m in Aceh⁸.

We sought geological evidence for past tsunamis on a marshy plain along 4 km of coast north of Meulaboh (Fig. 1a, Me). Marshes have previously yielded geological histories of tsunamis in the northwestern United States⁹, Kamchatka¹⁰, Japan¹¹ and Chile¹². In Aceh, the marshy plain shows a characteristic ridge and swale topography that is formed by rapid shoreline progradation. Individual beach ridges, which mark former positions of the shoreline, run parallel to the coast for several kilometres (Fig. 1b and Supplementary Fig. 1). They are separated by swales where peaty marsh deposits accumulate.

The sequence of ridges builds progressively seaward at a rate that can be determined from the ages of deposits on the beach ridge plain: the oldest deposit in a swale ~1,800 m distant from the current coastline is dated at between AD 780 and 990, or between 1,230 and

1,020 years ago (Fig. 2a, c). Assuming that at the time of its deposition the coastline was located at the foot of the beach ridge immediately before it—that is, at a distance of ~1,600 m from the current coastline—the average coastal progradation rate can be estimated to lie between 1.3 to 1.6 m yr⁻¹. Such a range has been described for similar beach ridge plains elsewhere^{13,14}.

In swales between beach ridges we took more than 100 auger cores along four transects running perpendicular from the coast up to 2 km inland (Fig. 1b and Supplementary Fig. 1). Wet ground and partly submerged areas limited access and precluded trenching (Supplementary Fig. 1). The coring in the swales was limited to depths of 1–2 m by compact grey-greenish, fine to medium shoreface sand extending beneath the beach ridge plain (Fig. 2 and

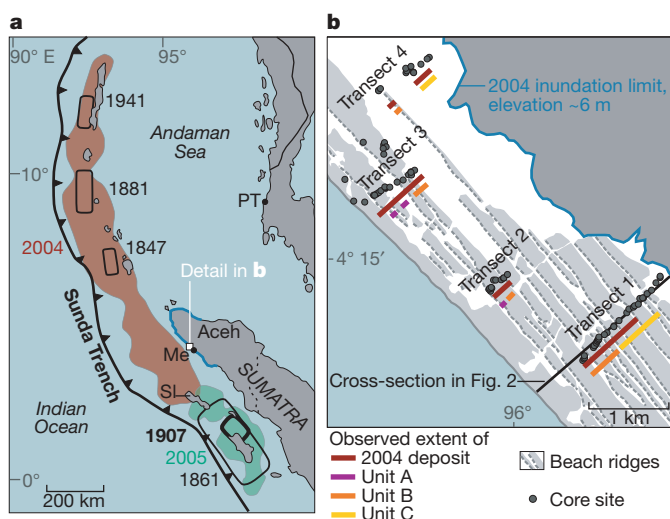


Figure 1 | Index maps. **a**, Northern Sunda margin and vicinity. The red patch denotes the area of seismic slip during the 2004 Sumatra–Andaman earthquake of magnitude 9.2 (ref. 26). The green patch indicates the area of seismic slip during the March 2005 earthquake⁷. Rectangular patches outlined in black mark inferred rupture areas of historically reported earthquakes^{3–6,26}. The blue line denotes Acehnese coast damaged by the 2004 tsunami. SI, Simeulue Island; Me, Meulaboh; PT, Phra Thong Island, Thailand. **b**, Locations of sample sites and extents of sand sheets A, B and C along shore-normal transects. The beach ridges, sketched from aerial photographs (Supplementary Fig. 1), mark former shorelines from the past 1,000 years or more. The 2004 tsunami inundation is mapped from the extent of dead trees in Supplementary Fig. 1.

¹Department of Geology, Kent State University, McGilvrey Hall, Kent, Ohio 44242, USA. ²Swiss Agency for Development and Cooperation (SDC), Swiss Humanitarian Aid Unit (SHA), Gerechtigkeitsgasse 20, 8002 Zürich, Switzerland. ³Old Woman Creek National Estuarine Research Reserve, 2514 Cleveland Road East, Huron, Ohio 44839, USA. ⁴Badan Pengkajian dan Penerapan Teknologi (BPPT), Jalan Grafika, Yogyakarta, 55281, Indonesia. ⁵Department of Earth Sciences and Geography, Vassar College, 124 Raymond Avenue, Poughkeepsie, New York 12604, USA. ⁶Department of Geosciences, Earlham College, 801 National Road West, Richmond, Indiana 47374, USA. ⁷Catholic Relief Services (CRS), Jalan Nasional, Meulaboh, NAD, Indonesia. [†]Present addresses: Department of Geology and Planetary Science, University of Pittsburgh at Johnstown, 450 Schoolhouse Road, Johnstown, Pennsylvania 15904, USA (K.M.); Franzius-Institute for Hydraulic, Waterways and Coastal Engineering, Leibniz Universität Hannover, Nienburger Strasse 4, 30167 Hannover, Germany (W.K.).

Supplementary Fig. 2). Because swale age increases inland, however, coring progressively farther from the modern beach allowed us to extend the sedimentary record about 1,000 years into the past.

North of Meulaboh the 2004 tsunami reached wave heights of between 9 and 14 m (ref. 15) and inundated the area to a distance of 2.0–2.5 km inland (Fig. 1b and Supplementary Fig. 1). The tsunami deposited a sand sheet that can be followed up to 1.8 km inland,

and shows overall landward thinning from ~50 cm near the shoreline to a few millimetres farther inland (Fig. 2 and Supplementary Fig. 2). A continuous sand sheet can be found in swales up to 800 m inland, whereas farther away from the shoreline the deposit becomes patchy. Although typically without obvious internal structure, in some locations the 2004 sand sheet shows normal grading with as many as three normally graded layers (Supplementary Fig. 3). Near the present shore the 2004 deposit consists of moderately-to-well-sorted medium sand dominated by silicate grains. Farther inland the sorting becomes poor; the mean grain size diminishes to fine sand (Supplementary Fig. 4), and the sand grains are mixed with soil fragments and plant debris. Microscopic fossils in 2004 deposits include pollen grains and diatoms. Although the diatoms are mostly of freshwater origin, one sample abounds in *Nitzschia salinicola*, a species that can grow in brackish water and is associated with sand grains (Supplementary Fig. 2 and Supplementary Table 1).

Sediment cores from the swales revealed three older sand beds, a few millimetres to 25 cm thick, that are intercalated with peaty marsh deposits (units A, B and C in Fig. 2, and Supplementary Fig. 2). We correlated them both between cores along shore-normal transects and between transects connected by shore-parallel beach ridges (Figs 1b and 2, and Supplementary Figs 1 and 4). Though nowhere encountered in vertical stratigraphic succession, the three units are superposed geomorphically: in keeping with the progressive seaward shift in shoreline, the oldest of the three units (C) is the farthest inland, and the youngest unit (A) the closest to the modern shoreline. The three units can be distinguished not just by geomorphic position on the beach ridge plain but also by landward fining along transects (Supplementary Fig. 4) and by radiocarbon ages (Supplementary Table 2).

Among the three pre-2004 sand sheets, unit A was found in just a few places 300–500 m from the modern beach, whereas units B and C extend more widely. Along transect 1 we found them 700–1,100 m inland and 1,200–1,800 m inland, respectively (Figs 1b and 2). All three units resemble the 2004 tsunami deposit compositionally: they are composed of grey-brown, silicate-rich, fine to medium sand rich in pollen grains and plant debris. Individual units become more fine-grained and less well sorted with increasing distance to the shoreline (Supplementary Fig. 4). Also like the 2004 deposit, the sand layers are typically massive but might show normal grading. Unit B contains two depositional units locally (Fig. 2d).

Among 14 microfossil samples, only 8 yielded diatoms (Supplementary Fig. 2). These are dominated by freshwater species (Supplementary Fig. 2 and Supplementary Table 1) except in one sample containing marine valves of *Actinopterychus* and *Diplomenora* (Fig. 2e, Supplementary Fig. 2 and Supplementary Table 1). We infer that the overall scarcity of diatoms in the pre-2004 sand units results from the dissolution of diatom silica at high temperatures and within an organic-rich environment¹⁶.

We obtained AMS radiocarbon ages on pieces of wood and other plant debris, all probably detrital, from below and above sand beds. We interpret the lower dates as maximum ages and the upper ones as minimum ages for the times of sand-sheet deposition (Supplementary Table 2). The resulting age estimates, with ranges at two standard deviations, are younger than 1640–1950 AD for

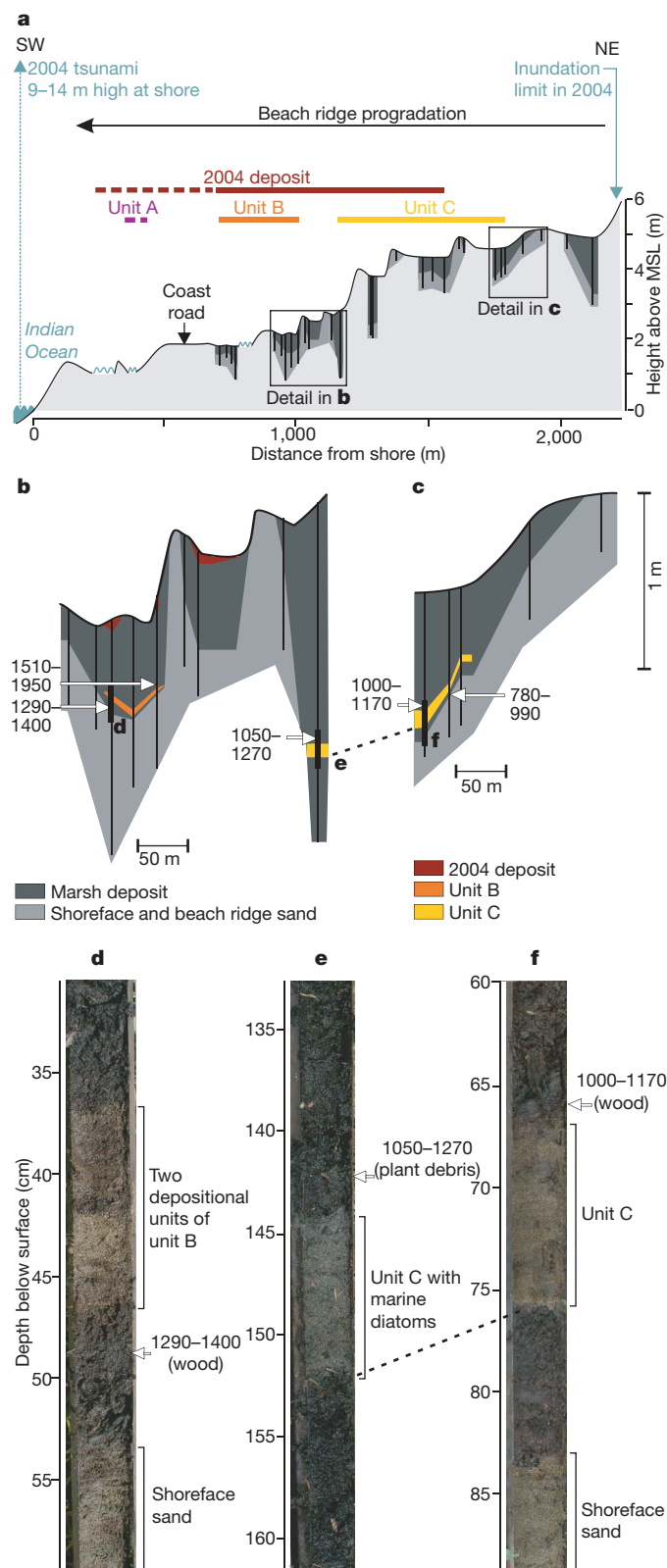


Figure 2 | Examples of sand sheets near Meulaboh. **a**, Topography of the beach ridge plain along shore-normal transect 1 (location shown in Fig. 1b), obtained by auto-level surveying, field observations, aerial photographs and a digital elevation model from helicopter surveys²⁷. Cores are shown by vertical lines. The observed extents of the 2004 sand sheet and of three earlier sheets (units A, B and C) are plotted above profile. Stippled lines mark the extent of sand sheets projected from near shore transects 2 and 3. The vertical elevation has been magnified 125-fold. MSL, mean sea level. **b**, **c**, Generalized stratigraphy and radiocarbon ages (in years AD at two standard deviations) along two parts of transect 1 indicated in **a**. **d**–**f**, Photographs of auger core samples from sand units B and C. Locations are shown in **b** and **c**.

unit A, between 1290–1400 and 1510–1950 AD for unit B, and between 780–990 and 1000–1170 AD for unit C.

Units A, B and C are best explained by tsunamis that ran onto the beach ridge plain and deposited sand on top of peaty marshes. Architecturally the three units resemble tsunami sand sheets described elsewhere: the sand extends hundreds of metres inland and rarely exceeds 25 cm in thickness¹⁷. Massive or normally graded beds and the lack of bed-load transport structures further indicate settling from suspension, which is characteristic of tsunami deposition^{17,18}. Finer-grained and less well sorted deposits with increasing distance from the shoreline have been observed in other tsunami deposits¹⁷ and point towards waning flow velocities and towards the incorporation of plant debris and soil into the flow during the landward passage of a tsunami. Despite the meagre microfossil evidence of a marine origin of the sand layers, we rule out flood deposition because most of our sampling sites are located a few kilometres south of the nearest (Woyla) river (Supplementary Fig. 1). Furthermore, villagers report that river floods during the monsoon season have not reached our sampling sites.

Sumatra's geographical setting probably excludes the deposition of the sand beds by storms. The island is located within the equatorial zone where tropical cyclones rarely occur for lack of sufficient Coriolis Force¹⁹. One exception was Typhoon Vamei in 2001, which formed in the South China Sea, proceeded westward and eventually passed over Sumatra, 100 km south of Meulaboh, before dissipating over the Indian Ocean²⁰. The storm weakened during its passage over Sumatra and caused no reported storm surge on the island's west coast²¹.

How might the three pre-2004 sand units correlate with the written and geological records in the Indian Ocean region? Unit A, being younger than AD 1640–1950 and limited to a narrow area, might represent the documented smaller tsunami of AD 1907. Unit B, if deposited shortly after AD 1290–1400, may correlate with the youngest pre-2004 tsunami deposit inferred from sand sheets of Phra Thong Island, Thailand (Fig. 1a), dated to soon after AD 1300–1450 (ref. 22). Unit C, deposited after AD 780–990, has no dated equivalent in Thailand. However, marine terraces in the Andaman Islands, the northern part of the 2004 rupture area (Fig. 1a), have been dated to AD 1170–1600 and AD 550–1330 and have been interpreted as evidence for subduction earthquakes²³. Acehnese units B and C might correlate with these terraces.

The combined evidence from Meulaboh and Phra Thong, although correlated tentatively, suggests that recurrence intervals of destructive tsunamis from Sumatra–Andaman sources can span centuries, with the 2004 Indian Ocean tsunami separated from its youngest full predecessor by perhaps 600 years. Such a long recurrence interval, which exceeds Indonesian tsunami records by about 200 years, would explain not just the lack of historical data but also the enormity of the 2004 Sumatra–Andaman earthquake²⁴.

Infrequent tsunami recurrence poses a dilemma for communities devastated by the 2004 tsunami. What balance should survivors seek between the risks of the next tsunami that might not return for a few generations, and the benefits of living close to sea? A partial solution is to ensure that communities sustain awareness of tsunami hazards. Such awareness saved thousands of lives on Simeulue Island (Fig. 1a), where people knew to interpret the severe ground shaking of the 2004 earthquake as a natural tsunami warning²⁵. This knowledge, derived from Simeulue's great losses to the 1907 tsunami, led the island's residents flee to high ground in time to escape the 2004 tsunami. On the Sumatran mainland, a large fraction of the tsunami's victims may have survived in this fashion had the area's history of recurring tsunamis been detected and communicated in time.

Received 7 March; accepted 27 August 2008.

1. Tsuji, Y. *et al.* Distribution of the tsunami heights of the 2004 Sumatra tsunami in Banda Aceh measured by the tsunami survey team (<http://www.eri.u-tokyo.ac.jp/namegaya/sumatera/surveylog/eindex.htm>) (2005).
2. Yalciner, A. C. *et al.* December 26, 2004 Indian Ocean tsunami field survey (Jan. 21–31, 2005) at north of Sumatra Island (<http://ioc.unesco.org/iosurveys/Indonesia/yalciner/yalciner.htm>) (2005).

3. Hamzah, L., Puspito, N. & Imamura, F. Tsunami catalogue and zones in Indonesia. *J. Nat. Disast. Sci.* **22**, 25–43 (2000).
4. Newcomb, K. R. & McCann, W. R. Seismic history and seismotectonics of the Sunda Arc. *J. Geophys. Res.* **92**, 421–439 (1987).
5. Bilham, R., Engdahl, R., Feldt, N. & Satyabala, S. P. Partial and complete rupture of the Indo-Andaman plate boundary 1847–2004. *Seismol. Res. Lett.* **76**, 299–311 (2005).
6. Ortiz, M. & Bilham, R. Source area and rupture parameters of the 31 December 1881 $M_w = 7.9$ Car Nicobar earthquake estimated from tsunamis recorded in the Bay of Bengal. *J. Geophys. Res.* **108**, 1–16 (2003).
7. Briggs, R. W. *et al.* Deformation and slip along the Sunda megathrust in the great 2005 Nias–Simeulue earthquake. *Science* **31**, 1897–1901 (2006).
8. U.S. Geological Survey. Notes From the Field ... USGS Scientists in Sumatra Studying Recent Tsunamis: Leg 2 Reports, 12 April to 30 April 2005 <<http://walrus.wr.usgs.gov/news/reports.html>> (2005).
9. Atwater, B. F. & Moore, A. L. A tsunami about 1000 years ago in Puget Sound, Washington. *Science* **258**, 1614–1617 (1992).
10. Pinegina, T. K. & Bourgeois, J. Historical and paleo-tsunami deposits on Kamchatka, Russia: long-term chronologies and long-distance correlations. *Nat. Hazards Earth Syst. Sci.* **1**, 177–185 (2001).
11. Nanayama, F. *et al.* Unusually large earthquakes inferred from tsunami deposits along the Kuril trench. *Nature* **424**, 660–663 (2003).
12. Cisternas, M. *et al.* Predecessors of the giant 1960 Chile earthquake. *Nature* **437**, 404–407 (2005).
13. Anthony, E. J. Beach-ridge development and sediment supply: examples from West Africa. *Mar. Geol.* **129**, 175–186 (1995).
14. Tanner, W. F. Origin of beach ridges and swales. *Mar. Geol.* **129**, 149–161 (1995).
15. Kongo, W., Istiyanto, D. C. & Irwandi, I. *Tsunami Modeling and Field Observations of December 26 2004 Indian Ocean Earthquake* (Technical Report, Coastal Dynamic Research Center (BPPT), Jogjakarta, 2006).
16. Bennett, P. C., Siegel, D. I., Hill, B. M. & Glaser, P. H. Fate of silicate minerals in a peat bog. *Geology* **19**, 328–331 (1991).
17. Morton, R. A., Gelfenbaum, G. & Jaffe, B. E. Physiscal criteria for distinguishing sandy tsunami and storm deposits using modern examples. *Sedim. Geol.* **200**, 184–207 (2007).
18. Goff, J., McFadgen, B. G. & Chagué-Goff, C. Sedimentary differences between the Easter storm and the 15th century Okoropunga tsunami, southeastern North Island, New Zealand. *Mar. Geol.* **204**, 235–250 (2004).
19. Anthes, R. A. *Tropical Cyclones: Their Evolution, Structure and Effects* (Am. Meteorol. Soc., 1982).
20. Chang, C.-P., Liu, C.-H. & Kuo, H.-C. Typhoon Vamei: An equatorial tropical cyclone formation. *Geophys. Res. Lett.* **30**, 1151–1154, doi:10.1029/2002GL016365 (2003).
21. Grandau, F. & Engel, G. 2001 Annual Tropical Cyclone Report. (U.S. Naval Pacific Meteorology and Oceanography Center/Joint Typhoon Warning Center, 2002).
22. Jankaew, K. *et al.* Medieval forewarning of the 2004 Indian Ocean tsunami in Thailand. *Nature* doi:10.1038/nature07373 (this issue).
23. Rajendran, K. *et al.* Age estimates of coastal terraces in the Andaman and Nicobar islands and their tectonic implications. *Tectonophysics* **455**, 53–60 (2008).
24. Stein, S. & Okal, E. A. Ultralong period seismic study of the December 2004 Indian Ocean earthquake and implications for regional tectonics and the subduction process; The 2004 Sumatra–Andaman earthquake and the Indian Ocean tsunami. *Bull. Seismol. Soc. Am.* **97**, 279–295 (2007).
25. McAdoo, B. G., Dengler, L., Prasetya, G. & Titov, V. Smong: How an oral history saved thousands on Indonesia's Simeulue Island during the December 2004 and March 2005 tsunamis. *Earthq. Spectra* **22**, 661–669 (2006).
26. Chlieh, M. *et al.* Coseismic slip and afterslip of the great M_w 9.15 Sumatra–Andaman earthquake of 2004. *Bull. Seismol. Soc. Am.* **97**, 152–173 (2007).
27. Siemon, B., Steuer, A., Voss, W. & Rehli, H. J. *Helicopter-Borne Geophysical Investigation in The Province of Nanggroe Aceh Darussalam, Northern Sumatra, Indonesia. Survey area Calang-Meulaboh* (Technical Report, Federal Institute for Geosciences and Natural Resources (BGR), Hanover, 2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the field offices of Caritas Switzerland and the Catholic Relief Services (CRS) in Meulaboh for logistical support during fieldwork; D. Plöthner and U. Meyer for providing maps and digital elevation data for the study area; the SIM data centre of the Aceh and Nias Rehabilitation and Reconstruction Board (BRR), in Banda Aceh, for providing aerial photographs; T. Ninck for help in the field; J. Beitel for assistance with laboratory work; K. Moran, S. Lüthi, S. Phipps, A. Smith, N. Wells and J. Ortiz for discussions; and B. Atwater, S. Bondevik, Y. Sawai and K. Sieh for reviews. The work was funded by the Swiss Science Foundation, the US National Science Foundation, and Vassar College.

Author Contributions K.M., W.F., W.K. and S.S. did the field work. D.K. performed the diatom analyses. A.M. and K.M. conducted the grain-size analyses. W.F. and B.M. did the mapping. K.M. prepared the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to K.M. (monecke@pitt.edu).

Agrochemicals increase trematode infections in a declining amphibian species

Jason R. Rohr^{1,2}, Anna M. Schotthoefer³, Thomas R. Raffel^{1,2}, Hunter J. Carrick⁴, Neal Halstead¹, Jason T. Hoverman⁵, Catherine M. Johnson⁶, Lucinda B. Johnson⁶, Camilla Lieske³, Marvin D. Piwoni⁷, Patrick K. Schoff⁶ & Val R. Beasley³

Global amphibian declines have often been attributed to disease^{1,2}, but ignorance of the relative importance and mode of action of potential drivers of infection has made it difficult to develop effective remediation. In a field study, here we show that the widely used herbicide, atrazine, was the best predictor (out of more than 240 plausible candidates) of the abundance of larval trematodes (parasitic flatworms) in the declining northern leopard frog *Rana pipiens*. The effects of atrazine were consistent across trematode taxa. The combination of atrazine and phosphate—principal agrochemicals in global corn and sorghum production—accounted for 74% of the variation in the abundance of these often debilitating larval trematodes (atrazine alone accounted for 51%). Analysis of field data supported a causal mechanism whereby both agrochemicals increase exposure and susceptibility to larval trematodes by augmenting snail intermediate hosts and suppressing amphibian immunity. A mesocosm experiment demonstrated that, relative to control tanks, atrazine tanks had immunosuppressed tadpoles, had significantly more attached algae and snails, and had tadpoles with elevated trematode loads, further supporting a causal relationship between atrazine and elevated trematode infections in amphibians. These results raise concerns about the role of atrazine and phosphate in amphibian declines, and illustrate the value of quantifying the relative importance of several possible drivers of disease risk while determining the mechanisms by which they facilitate disease emergence.

There is growing appreciation of the importance of infectious diseases in driving population dynamics and declines of wildlife species^{3,4}. However, investigators rarely assess the relative importance of the multitude of potential drivers and demonstrate a causal relationship between these factors and emerging infections so that the most problematic driver(s) could be effectively managed to prevent diseases of conservation concern⁵. Identifying the main risk factors for disease in amphibians is especially important because their widespread population declines have often been attributed to infectious disease^{1,2}. Trematode infections of amphibians have attracted research and public attention predominantly because they can cause grotesque limb malformations, kidney damage and debility^{6,7}. Furthermore, certain trematode infections are considered to be emerging diseases of amphibians driven by anthropogenic factors^{7–10}, and many can be lethal with the probability of amphibian mortality increasing with trematode load^{6,11,12}.

To understand the relative importance of drivers of amphibian trematode infections, we studied 18 wetlands in Minnesota, USA (Supplementary Table 1 and Supplementary Fig. 1), in which we

measured more than 240 plausible predictors of amphibian trematode infections (Supplementary Table 2). We quantified larval trematode abundance in northern leopard frogs (*R. pipiens*; Supplementary Fig. 2), a species in decline across much of its range¹³, and determined which of the factors best predicted the abundance of these parasites in these amphibians. The potential predictors we quantified included (1) the abundance of intermediate hosts of trematodes; (2) water quality and pollutants (Supplementary Table 3), which might suppress host immunity or be directly lethal to hosts or free-living stages of trematodes¹⁴; and (3) landscape attributes that might be crucial for supporting amphibians and avian and mammalian definitive hosts.

Our analyses showed that atrazine, the second most commonly used pesticide in the United States¹⁵ and perhaps the world, combined with one of its metabolites, desethylatrazine (both from water samples), was the best predictor of total larval trematode abundance in *R. pipiens*. The sum of atrazine and desethylatrazine accounted for more variation than any other single predictor (51%, Monte Carlo randomization test, $P = 0.001$, Fig. 1a and Supplementary Fig. 3) and occurred in more models than any other predictor in a best subset analysis (Supplementary Table 4). For subsequent analyses, we focused on total larval trematode abundance because of the similar responses of larval trematode taxa to atrazine (Supplementary Table 5), and on atrazine and desethylatrazine combined because the combination was a better predictor of larval trematode loads than either variable alone and because the two compounds have similar half-lives and toxicity¹⁶. Hereafter, for brevity, we refer to the sum of detectable atrazine and desethylatrazine concentrations in water as 'atrazine'.

We proposed that atrazine augmented the richness and abundance of gastropods, the first intermediate host of trematodes, increasing the abundance of free-living trematode cercariae and the consequent exposure of *R. pipiens* to trematodes (Supplementary Fig. 4)¹⁰. This hypothesis was on the basis of the observation that atrazine and other herbicides regularly cause net increases in periphytic algae¹⁷, a food source for many gastropods, despite herbicides being directly toxic to many periphyton species (see Supplementary Methods for mechanisms). In addition, we also suggested that atrazine increased the susceptibility of *R. pipiens* to trematode infections (Supplementary Fig. 4). This hypothesis was on the basis of laboratory studies which showed that low atrazine concentrations caused immunosuppression in ranid frog species^{18,19} and an ambystomatid salamander²⁰, increasing larval trematode loads and susceptibility to viral infections, respectively^{19,20}.

¹Biology Department, University of South Florida, Tampa, Florida 33620, USA. ²Penn State Center for Infectious Disease Dynamics, Penn State University, University Park, Pennsylvania 16802, USA. ³College of Veterinary Medicine, University of Illinois at Urbana-Champaign, Urbana, Illinois 61802, USA. ⁴School of Forest Resources, Penn State University, University Park, Pennsylvania 16802, USA. ⁵Department of Forest, Wildlife and Fisheries, The University of Tennessee, Knoxville, Tennessee 37996-4563, USA. ⁶Natural Resources Research Institute, University of Minnesota Duluth, Duluth, Minnesota 55811, USA. ⁷Illinois Waste Management and Research Center, Champaign, Illinois 61820, USA.

We used a path analysis to explore the amount of support for these proposed mechanisms for how atrazine might increase larval trematode loads²¹. Despite the low sample size of the path analysis, almost every path was significant (Fig. 1b) and the parameters were generally robust to bootstrapping. Parameter significance was also unaffected when controlling for the relationship between atrazine and phosphate (Supplementary Table 6), the latter of which is a primary ingredient in fertilizers that can also increase larval trematode loads¹⁰. The path model showed that atrazine was positively related to gastropod abundance and richness, and that gastropod richness was positively associated with larval trematode loads (Fig. 1b). Independent of this exposure-mediated effect on infection risk was a significant positive association between atrazine and larval trematode abundance (Fig. 1b), consistent with atrazine increasing the susceptibility of *R. pipiens* to trematodes.

To test this susceptibility hypothesis, we quantified melanomacrophage aggregates in the livers of frogs from each wetland (Supplementary Fig. 5). Melanomacrophages are cells of macrophage lineage (that is, Kupffer cells) found in the kidneys, spleen and liver that seem to be involved in inflammatory responses, in responding to foreign and dead material, and in fighting infections by a variety of

parasites including encysted larval helminths^{22–24}; furthermore, they have been proposed as reliable biomarkers for water pollution²⁴. We did not incorporate melanomacrophage aggregates into the path analysis because, unlike the other variables in the model (Fig. 1b), melanomacrophage aggregates seemed to be significantly increased by the richness and abundance of nematodes and adult trematodes that also infected the frogs (Supplementary Table 7). Consequently, we controlled for these co-infections when testing for relationships between melanomacrophage aggregates and atrazine, and between melanomacrophage aggregates and larval trematode loads. Hereafter, nematode- and adult trematode-independent melanomacrophage aggregate levels will be referred to as melanomacrophage aggregate scores.

We discovered that atrazine was a significant negative predictor of melanomacrophage aggregate scores in *R. pipiens* (Fig. 2a). Phosphate (in water samples) was also a significant negative predictor of melanomacrophage aggregate scores (Fig. 2b). Similarly, in a laboratory study, both atrazine and fertilizer decreased peripheral leukocyte concentrations in tiger salamanders²⁰. Moreover, the melanomacrophage aggregate score was a significant negative predictor of larval trematode abundance in *R. pipiens* (Fig. 2c), supporting the assertion that pollution-related immunosuppression was an important contributor to elevated larval trematode loads.

We found little evidence that atrazine was correlated with any other quantified variables that could parsimoniously explain the observed patterns; no variables that were correlated with atrazine and larval trematode loads were correlated with gastropod richness and melanomacrophage aggregate scores (Supplementary Table 8). Specifically, water column phosphate—which also can elevate amphibian trematode loads by increasing periphytic algae and gastropod abundance¹⁰—was not significantly correlated with atrazine and, when tested alone, was not a significant predictor of the abundance of larval trematodes in *R. pipiens* (Supplementary Table 8). However, when we included phosphate and atrazine in our statistical model, both variables were significant positive predictors of larval trematode burdens in *R. pipiens* ($F_{1,14} = 12.17$, $P = 0.004$, and $F_{1,14} = 31.06$, $P < 0.001$, respectively), indicating that they accounted for relatively uncorrelated components of variation in larval trematode loads in the study region. Moreover, the combination of atrazine and phosphate accounted for 74% of the variation in larval trematode abundance ($F_{2,14} = 19.75$, $P < 0.001$), and atrazine and phosphate were in 19 and 17 of the top 20 best subset models for larval trematode abundance, respectively (Supplementary Table 4).

We conducted a community-level mesocosm experiment to test whether the detected correlation between atrazine and amphibian trematode loads was indeed causal. Our mesocosm experiment showed that, in just one generation, cattle tanks with a single dose of atrazine at an expected environmental concentration had more than four times as many snails as control tanks (Table 1). Atrazine significantly reduced phytoplankton abundance, which resulted in significantly increased nutrient availability, water clarity and sunlight penetration to attached algae (Table 1). Tanks with atrazine had significantly greater periphyton levels than control tanks (Table 1), which presumably fuelled the elevated snail reproduction and perhaps increased snail richness in ponds with elevated atrazine (due to productivity–diversity relationships; Supplementary Methods).

As in the field survey, there was a significant association between atrazine exposure and elevated frog susceptibility in our mesocosm experiment. The mesocosm experiment was conducted in the southern range of *R. pipiens* where this species is declining¹³. Consequently, we conducted our mesocosm experiment on two congeners, *Rana palustris* (pickerel frog) and *Rana clamitans* (green frog). *Rana palustris* had significantly lower liver melanomacrophage counts in atrazine tanks compared to control tanks (Table 1), and a similar trend was observed for their liver eosinophils (Table 1), another important immune cell for fighting larval trematodes¹⁹. *Rana palustris* survival

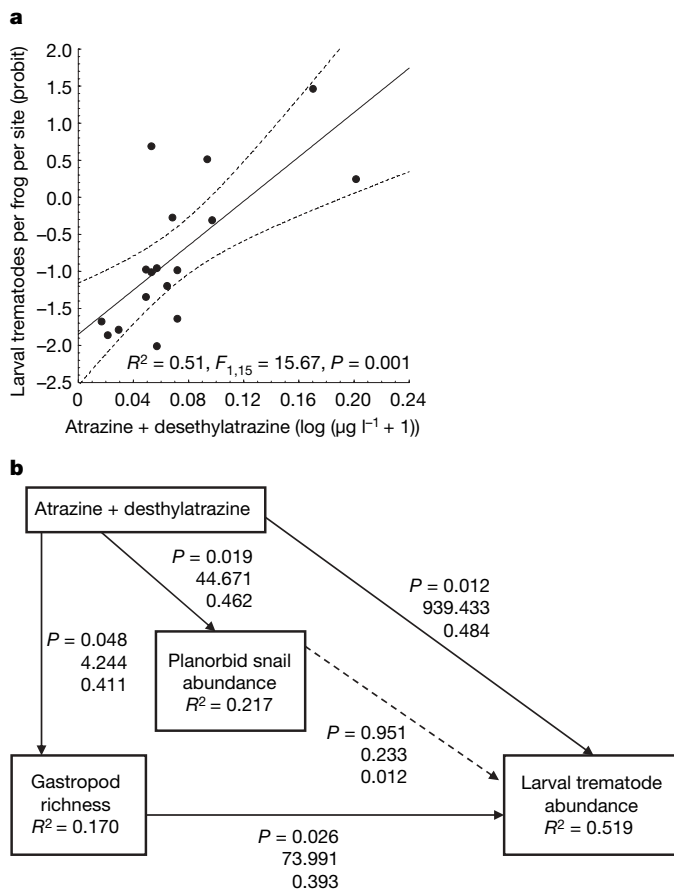


Figure 1 | Proposed mechanisms for the relationship between atrazine and larval trematodes in amphibians. **a**, The relationship between detectable concentrations of atrazine plus desethylatrazine in water samples and larval trematode abundance ('probit' transformed, that is, the normal standard deviate of mean abundance/1,000) in northern leopard frogs, *R. pipiens*. Each data point represents a wetland, and a 95% confidence band is shown. **b**, Results of a path analysis examining the suggested mechanisms by which atrazine plus desethylatrazine increase larval trematode abundance. Unstandardized and standardized coefficients, respectively, are provided next to each path. Solid arrows represent significant paths. The model chi-square was 0.339 (d.f. = 1, $P = 0.561$), indicating that the model was a good fit to the data.

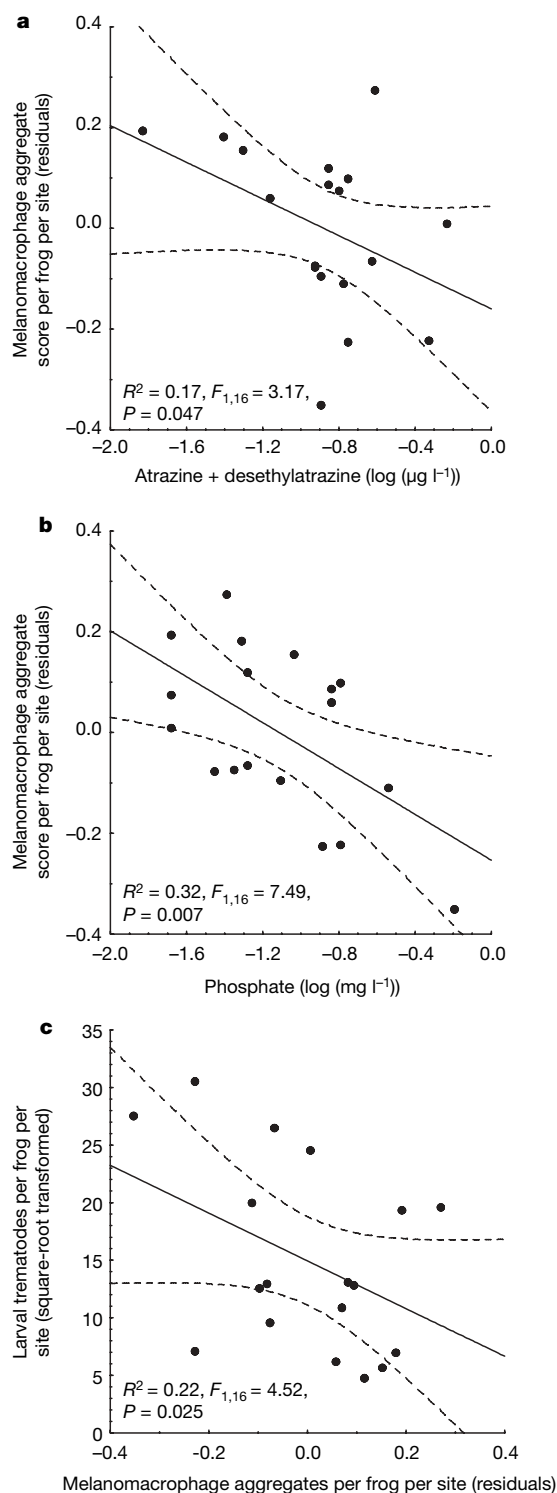


Figure 2 | Relationships between melanomacrophage aggregates and atrazine, phosphate and larval trematode loads. **a**, The relationship between melanomacrophage aggregate score in *R. pipiens* and concentrations of atrazine plus desethylatrazine in water samples. **b**, The relationship between melanomacrophage aggregate score in *R. pipiens* and phosphate concentration in water samples. **c**, The relationship between melanomacrophage aggregate score in *R. pipiens* and larval trematode abundance. In these graphs and analyses, we used the residuals from the relationship among mean melanomacrophage aggregate scores and the richness and abundance of nematodes and adult trematodes in leopard frogs (see text). Each data point represents a wetland, and 95% confidence bands are shown.

was significantly lower in the tanks with atrazine than in the tanks without atrazine (Table 1), preventing us from conducting an unconfounded test of the effect of atrazine on trematode loads in this species because it is probable that the most highly infected individuals were those that died.

For *R. clamitans*, atrazine exposure was not significantly associated with reduced liver melanomacrophages or elevated mortality, but it was associated with significant reductions in liver eosinophil counts (Table 1). Hence, in both ranid species, atrazine significantly suppressed immune parameters known to be important for fighting larval trematodes^{19,22–24}, a result consistent with previously documented immunosuppression in *R. pipiens* and *Rana sylvatica* exposed to atrazine^{18,19}. Furthermore, in atrazine tanks, *R. clamitans* had significantly more plagiophilid trematodes than in control tanks when controlling for the relationship between plagiophilid loads and eosinophils (Table 1), suggesting that atrazine had effects on trematode infections that were independent of its effects on eosinophils (Supplementary Methods).

Larval trematode loads in amphibians must depend on the abundance of definitive hosts, such as amphibian-eating birds, because they are the source of trematode eggs in wetlands; thus, an important caveat is that atrazine and phosphates cannot increase trematode infections unless there is an ample supply of trematodes to begin with. This dependence is supported by the significant interaction between atrazine and the number of habitat patches which were 'suitable' for definitive hosts (forest, wetlands and open water) around each wetland (Monte Carlo randomization test: 1 km radius, $P = 0.004$; 10 km radius, $P = 0.019$), the latter of which is a proxy for visitations by, and for the abundance of, definitive hosts. Knowledge of this context-dependency should facilitate appropriately targeting remediation strategies for amphibian trematode infections.

Our findings suggest that atrazine, at concentrations commonly occurring in freshwater ecosystems, can be a primary driver of larval trematode infections for a declining amphibian species, and that phosphate-associated eutrophication has an important, and perhaps complementary, role (because it was only significant when atrazine was included as a covariate). Both atrazine and phosphate appear to increase exposure and susceptibility to larval trematodes by augmenting snail intermediate hosts¹⁰ and by suppressing amphibian immune responses. By evaluating more than 240 alternative hypotheses, a number which could never realistically be examined in a manipulative experiment, and by demonstrating consistent results and underlying mechanisms between our controlled mesocosm experiment and wetland survey, we have substantially reduced the probability that the relationships identified are spurious. Hence, we provide support for the hypothesis that reducing atrazine and phosphate inputs to wetlands might effectively remediate elevated and often debilitating amphibian trematode infections.

Notably, the potentially important nexus between amphibian parasitism and pollution demonstrated here would not have been detected in standard studies used to register chemicals in the United States and Europe because these studies are typically conducted on individuals isolated from other species with which they naturally coexist, such as their parasites^{25,26}. Therefore, greater effort is needed to understand the potentially important effects of pollution on species interactions that may contribute to reduced host survival^{14,26}. Additionally, the results of this study—along with other studies linking pesticide use with amphibian declines²⁷ and linking low levels of atrazine with endocrine disruption, hermaphroditism²⁸ and mortality in amphibians^{29,30}—raise serious concerns about the relationship between atrazine and global amphibian losses. Altogether, our work demonstrates the value of quantifying the relative importance of several plausible drivers of disease risk and population declines using a combination of field surveys and manipulative studies to enable rational and prioritized environmental remediation.

Table 1 | Findings from mesocosm experiment examining the effect of atrazine on factors proposed to influence larval trematode loads in frogs.

Response variable	Atrazine (n = 4)		Control (n = 8)*		Test statistic	P†
	Mean	s.e.m.	Mean	s.e.m.		
Snail hatchlings m ⁻²	468.09	205.85	110.64	29.31	20.19‡	<0.001
Snail egg masses m ⁻²	30.45	3.03	12.37	4.34	4.98‡	0.026
Chlorophyll a in phytoplankton (µg l ⁻¹)	2.52	0.45	7.14	1.67	4.88§	0.031
Water clarity (scale 1–5)	4.75	0.25	3.88	0.23	5.25¶	0.011
Chlorophyll a in periphyton (mg m ⁻²)	1.04	0.18	0.70	0.08	7.17§	0.032
Melanomacrophages per field of view for <i>R. palustris</i> (log)	0.16	0.06	0.36	0.06	4.50§	0.031
Eosinophils per field of view for <i>R. palustris</i> (log)	0.03	0.14	0.21	0.06	1.88§	0.102 #
Survival of <i>R. palustris</i> (%)	47.50	7.50	70.00	3.90	9.77§	0.017
Melanomacrophages per field of view for <i>R. clamitans</i> (log)	0.25	0.13	0.38	0.09	0.67§	0.217 *
Eosinophils per field of view for <i>R. clamitans</i> (log)	0.11	0.10	0.39	0.07	4.75§	0.029
Survival of <i>R. clamitans</i> (%)	76.25	3.15	76.25	2.80	<0.001§	0.971**
No. of plagiogrichid trematode cysts in <i>R. clamitans</i>	20.61	4.52	7.51	3.30	4.90§	0.031 ††

* The solvent and water controls were pooled because there was no significant difference between these treatments.

† Block was included in the model if its *P* value was less than 0.25.

‡ χ^2 statistic from analyses using the generalized linear model with a Poisson error distribution and a log link.

§ *F* statistic from analyses using the general linear model.

|| One-tailed test.

¶ χ^2 statistic from analyses using the generalized linear model with an ordinal multinomial error distribution and a logit link.

The power of the test was 0.646; the minimum sample size necessary per group to detect a significant difference given the power was six.

* The power of the test was 0.326.

** The power of test was 0.056.

†† Controlling for biomass of plagiogrichid-infected snails ($F_{1,7} = 5.45$, $P = 0.026$) and mean log eosinophil counts in tadpoles ($F_{1,7} = 3.58$, $P = 0.050$).

METHODS SUMMARY

In each wetland in 1999, we characterized biological diversity (three visits: March–April, May–June and July–August; Supplementary Tables 2 and 3) and obtained water, sediment and amphibian tissue samples to quantify various analytes (two visits: April–May and June–July; Supplementary Table 3). Analyte quantification was conducted by the Illinois Waste Management and Research Center using standard US Environmental Protection Agency protocols. During wetland visits, we attempted to collect 15 recently metamorphosed *R. pipiens* for parasite quantification and 25 for pathology.

Our mesocosm experiment was conducted at Pennsylvania State University using a randomized block design and twelve 1,100-litre plastic cattle tanks (800 l of water) covered with shade cloth. Three weeks before dosing, each tank received 300 g of mixed hardwood leaves, plexiglass to sample snail egg masses and hatchlings, periphyton samplers and inoculations of zooplankton, periphyton and phytoplankton from four ponds. Just before dosing, each tank received four larval amphibian (5 *Ambystoma maculatum*, 20 *Hyla versicolor*, 20 *R. palustris* and 20 *R. clamitans*), two snail (11 *Planorbella trivolvis* and 10 *Physa gyrina*), one beetle (5 *Hydrochara* sp.), two water bug (2 *Belostomatidae* and 7 *Notonecta undulata*) and one dragonfly species (2 *Anax junius*). We applied a single dose of technical grade atrazine at 102 µg l⁻¹, an expected environmental concentration (mean actual concentration was 117 µg l⁻¹). The experiment ran for 4 weeks, from the end of June to July 2007.

All variables were quantified blind to the wetland characteristics or mesocosm treatment. Wetlands and mesocosms were used as the sample units for statistical analyses. That is, if several measurements were taken on a wetland or a tank, statistical analyses were conducted on the mean of those measurements. The path analysis was conducted assuming linear relationships among the raw data and using maximum likelihood estimation and bootstrapping²¹.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 15 April; accepted 24 July 2008.

- Stuart, S. N. *et al.* Status and trends of amphibian declines and extinctions worldwide. *Science* **306**, 1783–1786 (2004).
- Daszak, P., Cunningham, A. A. & Hyatt, A. D. Infectious disease and amphibian population declines. *Divers. Distrib.* **9**, 141–150 (2003).
- Hudson, P. J., Dobson, A. P. & Newborn, D. Prevention of population cycles by parasite removal. *Science* **282**, 2256–2258 (1998).
- de Castro, F. & Bolker, B. Mechanisms of disease-induced extinction. *Ecol. Lett.* **8**, 117–126 (2005).
- Lafferty, K. D. & Gerber, L. R. Good medicine for conservation biology: The intersection of epidemiology and conservation theory. *Conserv. Biol.* **16**, 593–604 (2002).
- Schotthoefer, A. M., Cole, R. A. & Beasley, V. R. Relationship of tadpole stage to location of echinostome cercariae encystment and the consequences for tadpole survival. *J. Parasitol.* **89**, 475–482 (2003).

- Johnson, P. T. J. & Sutherland, D. R. Amphibian deformities and *Ribeiroia* infection: An emerging helminthiasis. *Trends Parasitol.* **19**, 332–335 (2003).
- Skelly, D. K. *et al.* in *Disease Ecology: Community Structure and Pathogen Dynamics* (eds Collinge, S. K. & Ray, C.) 153–167 (Oxford Univ. Press, 2006).
- Beasley, V. R. *et al.* in *Status and Conservation of U.S. Amphibians* (ed. Lannoo, M. J.) 153–167 (Univ. Chicago Press, 2003).
- Johnson, P. T. J. *et al.* Aquatic eutrophication promotes pathogenic infection in amphibians. *Proc. Natl Acad. Sci. USA* **104**, 15781–15786 (2007).
- Johnson, P. T. J., Lunde, K. B., Ritchie, E. G. & Launer, A. E. The effect of trematode infection on amphibian limb development and survivorship. *Science* **284**, 802–804 (1999).
- Schotthoefer, A. M., Koehler, A. V., Meteyer, C. U. & Cole, R. A. Influence of *Ribeiroia ondatrae* (Trematoda: Digenea) infection on limb development and survival of northern leopard frogs (*Rana pipiens*): effects of host stage and parasite-exposure level. *Can. J. Zool.* **81**, 1144–1153 (2003).
- Rorabaugh, J. C. in *Amphibian Declines: The Conservation and Status of United States Species* (ed. Lannoo, M. J.) (Univ. California Press, 2005).
- Rohr, J. R., Raffel, T. R., Sessions, S. K. & Hudson, P. J. Understanding the net effects of pesticides on amphibian trematode infections. *Ecol. Appl.* (in the press).
- Kiely, T., Donaldson, D. & Grube, A. *Pesticide Industry Sales and Usage: 2000 and 2001 Market Estimates* (U.S. Environmental Protection Agency, 2004).
- Gammon, D. W. *Atrazine Risk Characterization Document* (California Department of Pesticide Regulation, 2001).
- Brock, T. C. M., Lahr, J. & Van den Brink, P. J. *Ecological Risks of Pesticides in Freshwater Ecosystems Part 1: Herbicides* Alterra-Rapport 088 (Alterra, Green World Research, 2000).
- Brodtkin, M. A., Madhoun, H., Rameswaran, M. & Vatnick, I. Atrazine is an immune disruptor in adult northern leopard frogs (*Rana pipiens*). *Environ. Toxicol. Chem.* **26**, 80–84 (2007).
- Kiesecker, J. M. Synergism between trematode infection and pesticide exposure: A link to amphibian limb deformities in nature? *Proc. Natl Acad. Sci. USA* **99**, 9900–9904 (2002).
- Forson, D. & Storfer, A. Effects of atrazine and iridovirus infection on survival and life-history traits of the long-toed salamander (*Ambystoma macrodactylum*). *Environ. Toxicol. Chem.* **25**, 168–173 (2006).
- Grace, J. B. *Structural Equation Modeling and Natural Systems* (Cambridge Univ. Press, 2006).
- Dezfuli, B. S. *et al.* Histopathology and ultrastructure of *Platichthys flesus* naturally infected with *Anisakis simplex* S.L. larvae (Nematoda: Anisakidae). *J. Parasitol.* **93**, 1416–1423 (2007).
- Reyes, J. L. & Terrazas, L. I. The divergent roles of alternatively activated macrophages in helminthic infections. *Parasite Immunol.* **29**, 609–619 (2007).
- Agius, C. & Roberts, R. J. Melano-macrophage centres and their role in fish pathology. *J. Fish Dis.* **26**, 499–509 (2003).
- Rohr, J. R. & Crumrine, P. W. Effects of an herbicide and an insecticide on pond community structure and processes. *Ecol. Appl.* **15**, 1135–1147 (2005).
- Rohr, J. R., Kerby, J. L. & Sih, A. Community ecology as a framework for predicting contaminant effects. *Trends Ecol. Evol.* **21**, 606–613 (2006).
- Davidson, C., Shaffer, H. B. & Jennings, M. R. Declines of the California red-legged frog: Climate, UV-B, habitat, and pesticides hypotheses. *Ecol. Appl.* **11**, 464–479 (2001).

28. Hayes, T. B. *et al.* Hermaphroditic, demasculinized frogs after exposure to the herbicide atrazine at low ecologically relevant doses. *Proc. Natl Acad. Sci. USA* **99**, 5476–5480 (2002).
29. Storrs, S. I. & Kiesecker, J. M. Survivorship patterns of larval amphibians exposed to low concentrations of atrazine. *Environ. Health Perspect.* **112**, 1054–1057 (2004).
30. Rohr, J. R., Sager, T., Sesterhenn, T. M. & Palmer, B. D. Exposure, postexposure, and density-mediated effects of atrazine on amphibians: Breaking down net effects into their parts. *Environ. Health Perspect.* **114**, 46–50 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank J. Murphy, A. Antolin, K. Beckman, R. Cole, A. Koehler, C. Hall, T. Hollenhorst and J. Romansic for collection and compilation of field, parasitological or land cover data; R. Cole, M. Martin, M. Mescher, J. Romansic, J. Runyon and the O. Bjørnstad, C. De Moraes, E. Holmes, P. Hudson, B. Grenfell and M. Poss laboratories for comments and suggestions on this work; and J. Grace for reviewing sections of the paper on structural equation modelling. We also thank the USGS National Wildlife Health Center for laboratory space and

support. Funds came from National Science Foundation (DEB-0809487) and US Department of Agriculture (NRI 2008-00622 and 2008-01785) grants to J.R.R., and US Environmental Protection Agency STAR grants to V.R.B. (R825867) and J.R.R. and T.R.R. (R833835). This work does not necessarily reflect the views of these agencies.

Author Contributions For the field survey, V.R.B. and L.B.J. designed the data collection. C.M.J., P.K.S., C.L. and A.M.S. conducted the survey. C.M.J. coordinated data collection, assembly and management. M.D.P. conducted all analyte analyses. A.M.S. performed amphibian necropsies of *R. pipiens* for parasite quantification. C.L. quantified amphibian immunity. For the mesocosm study, J.R.R., T.R.R. and J.T.H. designed and implemented the experiment. J.R.R. oversaw all components the study. T.R.R. and N.H. processed amphibian samples and quantified amphibian immune parameters. H.J.C. quantified periphyton and phytoplankton. J.R.R. conducted all statistical analyses and wrote the paper. A.M.S. wrote parts of the Supplementary Methods. The paper was edited by all authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to J.R.R. (jasonrohr@gmail.com).

METHODS

Wetland survey. Quantification of faunal diversity (amphibians and macroinvertebrates) was conducted using daytime visual time-constrained and area-constrained dip-net sampling methods, adjusting effort according to the size of the wetland. Three 10-m transects were randomly placed within each community type in and around each wetland, and the line-intercept method was used to record the relative cover of each plant species under or over the line.

Landscape structure—including landscape composition, fragmentation patterns, patch density and connectivity patterns—was quantified around each wetland using Fragstats (version 2). Land cover data were derived from the 1990s National Land Cover Database, which was based primarily on 1992 Landsat-5 Thematic Mapper data, with a 30-m resolution. These data were used to summarize landscapes at 1-km and 10-km extents. The 1-km extent was chosen because it represents the approximate dispersal capabilities of leopard frogs, and it allowed for exploration of effects of adjacent land use on the parasite communities. The 10-km scale was examined because this larger spatial scale might affect the highly mobile definitive hosts of many frog parasites (for example, birds and mammals).

The buccal and abdominal cavities were examined, and all visceral organs, skin and eyes were removed from necropsied frogs and inspected for parasites. After necropsy, the remaining frog carcasses were fixed in 10% neutral buffered formalin and cleared and stained. Encysted metacercariae in the musculature were identified and enumerated by examining the cleared and stained specimens under a dissecting microscope (Supplementary Fig. 2).

Frog livers were embedded in paraffin, stained with haematoxylin and eosin, sectioned and then examined under a light microscope. Melanomacrophage aggregate numbers were scored from zero to five, in which zero indicated that no melanomacrophage aggregates were present and five indicated that more than 90% of the liver was filled with melanomacrophage aggregates (Supplementary Fig. 5). Livers that fell in between categories were assigned intermediate values. All quantification was done on frogs of Gosner stage 46.

Our statistical analysis for the wetland survey was done in four steps. In the first step, we calculated the coefficient of determination (R^2) for each predictor in the absence of covariates to determine which single predictor accounted for the most variation in larval trematode abundance. In our second step, we used the general linear model to generate the 20 best subset models for mean larval trematode loads in *R. pipiens*. Models were ranked by adjusted R^2 values and were limited to a maximum of three predictors (Supplementary Table 4). In our third step, we used a structural equation model, which was later reduced to a path analysis on the basis of a confirmatory factor analysis (see Supplementary Methods and Supplementary Fig. 6), to evaluate the amount of support for the specific proposed causal mechanisms underlying the significant positive relationship between atrazine concentration and larval trematode abundance (Fig. 1b). Phosphate was not included in our path model because the primary emphasis was on atrazine and our sample size was too small to include both chemicals in the model. Path analysis can be sensitive to relatively small samples

sizes, raising concerns about the precision of parameter estimates and the stability of our path model. To evaluate the robustness of the parameter estimates and model stability, we obtained standard error estimates for all the paths in Fig. 1b using both maximum likelihood and bootstrapping methods (200 resamplings)²¹. Another matter is that herbicides and fertilizers might positively co-vary in the landscape, raising concerns about the validity of the path model if the relationship between phosphate concentration and atrazine was not controlled. We re-calculated all the parameter estimates and standard errors controlling for the relationship between atrazine and phosphate concentrations. Structural equation modelling and path analysis and bootstrapping were done using SEPATH in Statistica 6.1 (StatSoft Inc.). The final step was to assess whether atrazine was correlated with any other variables that might more parsimoniously explain the observed patterns. We used correlation analysis to identify any variables that were positively correlated with atrazine, larval trematode loads, gastropod richness and melanomacrophage aggregate scores.

Mesocosm experiment. All of the species in the mesocosm experiment naturally coexist and were applied at densities found locally. The applied expected environmental concentration of atrazine was calculated using the US Environmental Protection Agency's GENEEC (version 2) software (Supplementary Methods).

All snails used in the experiment were screened twice for patent (shedding) trematode infections before being placed in the tanks. In addition to the 11 screened planorbid snails that were free to move in each tank, each tank also received a caged planorbid snail shedding cercariae from the family Plagiorchidae. These caged snails were rotated among the tanks every other day in an attempt to homogenize exposure to trematodes. Snails were thoroughly rinsed before being moved to another tank and the cages were not rotated.

At the end of the experiment, we took water samples for phytoplankton analyses, scraped periphyton from the periphyton samplers, scored water clarity on a scale from 1 to 5, and counted the snail egg masses on the walls of each tank, the snail hatchlings from the plexiglass placed in each tank and the surviving and dead animals. Chlorophyll *a* concentrations were measured using standard fluorometric techniques. Five preserved *R. clamitans* and five *R. palustris* tadpoles from each tank were selected randomly and melanomacrophage and eosinophils cells were counted per field of view from their sectioned and stained livers. Each tadpole was also cleared and stained and their larval trematodes were counted under a compound scope, as described previously.

We re-screened each snail from trematodes at the end of the experiment and discovered that some of the snails that were placed in the tanks were pre-patent (infected but not shedding trematodes) and began shedding plagiorchid cercariae during the experiment. Consequently, all snails were weighed and dissected to ensure that we did not miss any infected snails that were not shedding at the end of the experiment. One tank had a snail infected with strigeid trematodes (determined on the basis of snail dissection and shedding) and we could not discriminate these trematode metacercariae from the plagiorchid metacercariae, and thus, this tank had to be dropped from the trematode load analyses.

LETTERS

Small-amplitude cycles emerge from stage-structured interactions in *Daphnia*–algal systems

Edward McCauley¹, William A. Nelson² & Roger M. Nisbet³

A long-standing issue in ecology is reconciling the apparent stability of many populations with robust predictions of large-amplitude population cycles from general theory on consumer–resource interactions¹. Even when consumers are decoupled from dynamic resources, large-amplitude cycles can theoretically emerge from delayed feedback processes found in many consumers^{2,3}. Here we show that resource-dependent mortality and a dynamic developmental delay in consumers produces a new type of small-amplitude cycle that coexists with large-amplitude fluctuations in coupled consumer–resource systems. A distinctive characteristic of the small-amplitude cycles is slow juvenile development for consumers, leading to a developmental delay that is longer than the cycle period. By contrast, the period exceeds the delay in large-amplitude cycles. These theoretical predictions may explain previous empirical results on coexisting attractors found in *Daphnia*–algal systems^{4,5}. To test this, we used bioassay experiments that measure the growth rates of individuals in populations exhibiting each type of cycle. The results were consistent with predictions. Together, the new theory and experiments establish that two very general features of consumers—a resource-dependent juvenile stage duration and resource-dependent mortality—combine to produce small-amplitude resource–consumer cycles. This phenomenon may contribute to the prevalence of small-amplitude fluctuations in many other consumer–resource populations^{6,7}.

Incorporating size- or age-structured interactions into theory on the dynamics of biological populations has provided new insight into the causes of population fluctuations and patterns in natural systems. Theory on single-species dynamics reveals that interactions among life-history stages or age classes and the resultant feedback on demographic rates^{2,8} can generate a diversity of cycle types. Many of these findings carry over to situations with indirect interactions among stages, for example through competition for food.

The range of possible mechanisms for explaining cyclic dynamics expands markedly once models include explicit interactions with other species. For example, stage-structure and explicit consumer–resource interactions can each destabilize an equilibrium, leading to population cycles^{3,9–11}. Intuitively one might think that combining these sources of instability would enhance instability in dynamics. Here, we demonstrate that this combination does not simply increase the amplitude of cycles, but introduces new phenomena that arise from the interaction of the various sources of instability. We show that combining one of the most general life histories for a predator with consumer–resource dynamics produces new types of cycles that may help to explain the rarity of large-amplitude cycles in many systems.

Our approach combines theory on the dynamics of stage-structured populations of the herbivore *Daphnia pulex* and their algal prey with experiments conducted in microcosms to test mechanisms producing different types of cycles. By coupling small-volume

flow-through systems to microcosms containing the *Daphnia*–algal populations, we measure the growth rate of individual herbivores experiencing food fluctuations set by the dynamic interactions unfolding in the microcosms, and compare these estimates to model predictions for different types of cycles.

Previous work^{3,11,12} has identified three types of cycles that can emerge from stage-structured competition for food in a population being reared under semi-chemostat conditions. Which cycles are found depends on the relative feeding rates of juveniles and adults, fecundity, and relative ability to withstand starvation of juveniles and adults. Here we add to the previous models two general features of natural consumer–resource systems: density-dependent dynamics of resources and a saturating functional response for the consumers. This broadens the range of potential outcomes, including previously undiscovered types of cycle, and defines new empirical tests using consumer–resource systems, reported below. Our analysis exploits a new numerical bifurcation approach for stage-dependent delayed integro-differential models that facilitates resolution of regions of hysteresis, and allows us to identify robust properties of the different types of cycles that emerge.

Our stage-structured model for resource–consumer dynamics (Box 1 and Supplementary Information) is motivated by the extensive work on dynamics of *Daphnia* populations interacting with algal prey. Juvenile and adult predators (stages of *Daphnia*) differ in their feeding rates, energy allocation patterns, and resource-dependent mortality. We assume that the prey population is unstructured and prey cannot evolve. The resulting model that describes interactions of a ‘structured’ predator population with a simple prey has received relatively little previous attention¹³, in contrast with a large body of theory that includes ‘structure’ in the prey or resource^{14–16}, used in studies of host–parasitoid interactions or disease dynamics². Yet it contains the key features of many systems where prey complexity can be controlled.

Empirically, the *Daphnia*–algal system displays an extraordinary range of dynamics that includes fluctuations around a stable equilibrium, and a combination of small- and large-amplitude cycles^{4,6,7,17}. As predicted by previous theory, large-amplitude consumer–resource cycles can be found under nutrient-rich conditions, the so-called ‘paradox of enrichment’⁴. However, in these same environmental conditions, small-amplitude cycles not predicted by previous theory were present and there is evidence that these periodic attractors coexist^{4,5}. This suggests that there is some fundamental feature of this system that produces these small-amplitude cycles even under nutrient-rich conditions where only large-amplitude cycles are expected.

Bifurcation analysis of the stage-structured predator–prey model parameterized for the *Daphnia*–algal system (Fig. 1 and Supplementary Information) reveals new dynamics as the algal carrying capacity is increased. In environments with low algal carrying

¹Ecology and Evolution Group, Department of Biological Sciences, University of Calgary, Calgary, Alberta T2N 1N4, Canada. ²Department of Biology, Queen's University, Kingston, Ontario K7L 3N6, Canada. ³Department of Ecology, Evolution, and Marine Biology, University of California, Santa Barbara, California 93106, USA.

Box 1 | A structured model of the dynamics of a consumer and its resource

Population balance equations are shown for processes and interactions, along with equations for juvenile development and survival. The top three equations describe the dynamics of resource (F) and juvenile (J) and adult (A) consumers, respectively. The dynamics of juvenile development time and juvenile survival are determined by integral equations (bottom two equations). Further model details, parameter definitions and estimation, and bifurcation analysis is provided in Supplementary Information.

Resource dynamics:

$$\frac{dF(t)}{dt} = \overbrace{qF(t)\left(1 - \frac{F(t)}{k}\right)}^{\text{logistic resource growth}} - \overbrace{I_J \frac{F(t)}{F(t) + f_h} J(t)}^{\text{ingestion by juveniles}} - \overbrace{I_A \frac{F(t)}{F(t) + f_h} A(t)}^{\text{ingestion by adults}}$$

Juvenile dynamics:

$$\begin{aligned} \frac{dJ(t)}{dt} = & \overbrace{\frac{\chi I_A \sigma_A}{\gamma} \frac{F(t)}{F(t) + f_h} A(t)}^{\text{birth}} - \overbrace{\frac{\chi I_A \sigma_A}{\gamma} \frac{F(t)}{F(t) + f_h} A(t - \tau(t)) S(t)}^{\text{maturation}} \\ & - \overbrace{\frac{\mu_J}{I_J \sigma_J} \frac{F(t) + f_h}{F(t)} J(t)}^{\text{juvenile mortality}} \end{aligned}$$

Adult dynamics:

$$\begin{aligned} \frac{dA(t)}{dt} = & \overbrace{\frac{\chi I_A \sigma_A}{\gamma} \frac{F(t)}{F(t) + f_h} A(t - \tau(t)) S(t)}^{\text{maturation}} - \overbrace{\frac{\mu_A}{I_A \sigma_A} \frac{F(t) + f_h}{F(t)} A(t)}^{\text{adult mortality}} \\ \frac{w}{\sigma_J I_J} = & \int_{t-\tau(t)}^t \frac{F(\xi)}{F(\xi) + f_h} d\xi \quad \text{(juvenile development)} \\ S(t) = & \exp\left(-\frac{\mu_J}{I_J \sigma_J} \int_{t-\tau(t)}^t \frac{F(\xi) + f_h}{F(\xi)} d\xi\right) \quad \text{(juvenile survival)} \end{aligned}$$

capacity the model has a stable equilibrium, as is found in unstructured predator–prey models. Increasing algal carrying capacity beyond $\sim 0.3 \text{ mg C l}^{-1}$, we encounter a bifurcation near which two cycles appear: a new small-amplitude resource–consumer cycle and, via a subcritical fold of periodic orbits, a large-amplitude resource–consumer cycle. Notably, the two cycles are coexisting attractors over a broad range of algal carrying capacities corresponding to nutrient-rich environments (Fig. 1a). Bifurcation analysis also reveals that the cycles are separated by an unstable limit cycle. This causes the transition between the large- and small-amplitude cycles to occur abruptly (Supplementary Information), which is a common feature of laboratory and field experiments.

The two types of cycles can be distinguished by a key diagnostic feature^{18,19}: the ratio of the cycle period to juvenile stage duration (hereafter referred to as the delay). This is illustrated in Fig. 1b–e. In the large-amplitude cycles (Fig. 1b, c), the period exceeds the delay, a property that is characteristic of previously studied resource–consumer systems. By contrast, in the small-amplitude cycle (Fig. 1d, e), the juvenile stage duration is always larger than the cycle period, yielding cycles with a period/delay ratio less than one. The small-amplitude cycles are caused by the interaction between the instability caused by the resource–consumer interaction and the resource-dependent juvenile stage duration in the consumer *Daphnia*. Numerical experiments (not shown) established that removing the type II functional response that is the source of the instability in nutrient-rich algal environments, decoupling the prey dynamics by replacing density-dependent growth with semi-chemostat dynamics,

or fixing the juvenile stage duration at time periods longer than the cycle period, all cause the coexisting attractors to disappear.

Our structured model thus apparently explains one of the major empirical results reported previously⁴: the *Daphnia*–algal system can display both large- and small-amplitude cycles under the same nutrient-rich environment. This hypothesis can be rigorously tested by measuring the period/delay ratio in dynamic interacting populations. There are some major challenges in measuring the development time of individuals (stage durations) non-destructively in a population where individuals cannot be marked to follow their growth or performance. Some attempts to dye individual *Daphnia* have been tried, but feeding performance is affected. We overcame these challenges by designing new bioassay experiments whereby we can measure growth of individuals in populations experiencing food dynamics set by the interaction between predators and prey (Fig. 2 and Supplementary Information). Estimates of growth rates from these bioassays were validated through comparisons with ‘artificial cohort’ experiments (Supplementary Information).

Replicate *Daphnia*–algal systems display the same collection of dynamics described previously⁴. Under common environmental conditions of nutrient enrichment, temperature and light levels, the predator–prey systems display either large-amplitude or small-amplitude cycles, with transitions between the attractors (Fig. 2 and Supplementary Information). Consistent with dynamics presented previously⁴, small-amplitude cycles predominate ($\sim 80\%$ of the replicates) when resting egg production is not manipulated. Large- and small-amplitude cycles can be distinguished using two probes: variation in *Daphnia* egg density and in algal biomass (Fig. 3 and Supplementary Information). The cycle periods for these dynamics average 21.4 days with lower and upper 95% prediction intervals of 14.5 and 28.2, respectively (Fig. 4). The bioassay experiments

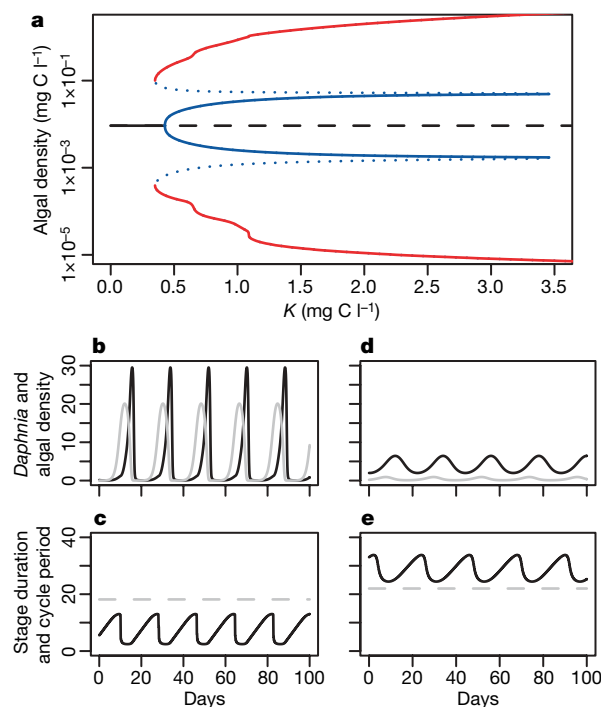


Figure 1 | Multiple limit-cycle attractors in the structured predator–prey model. **a**, Bifurcation diagram showing the transition from a stable steady state (solid black line) to a region of multiple coexisting limit cycles with increasing algal carrying capacity K (mg C l^{-1}). The range in algal density (mg C l^{-1}) over a cycle is shown. Stable small-amplitude cycles (blue) and large-amplitude cycles (red) are shown, separated by an unstable cycle (dashed blue line). **b**, **d**, Large- and small-amplitude cycles of *Daphnia* (black) and algae (grey). **c**, **e**, Graphs showing a key diagnostic feature: the relationship between cycle period (dashed line) and the stage duration of *Daphnia* (solid line) during large- (**c**) and small-amplitude cycles (**e**).

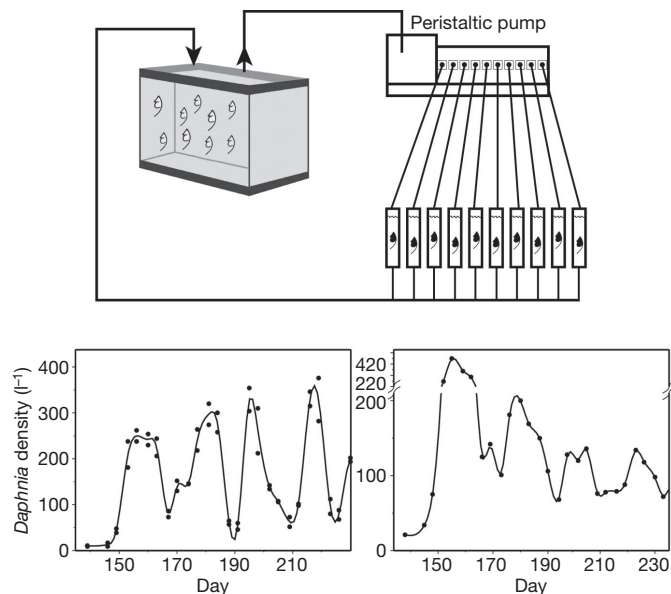


Figure 2 | Bioassay experiments for resource-consumer systems.

Distinguishing among causes of dynamics often requires observations on the performance of individuals in interacting populations. For many populations, individuals cannot be marked or followed, which restricts the estimation of key rates. Using flow-through systems coupled to microcosms (upper panel), we can estimate these rates for individuals. As the dynamics unfold in the microcosm, water containing algal prey flows through the vial housing an individual whose growth and reproduction can be observed during cyclic dynamics. Lower panels illustrate examples of large- and small-amplitude dynamics from which individuals were drawn to estimate length-specific growth rates.

(Fig. 2) show that individual *Daphnia* grow at different rates in the large-amplitude and small-amplitude cycles (Fig. 4a,b), and thus the development time bears a fundamentally different relationship with the cycle period. The juvenile stage duration of individual *Daphnia* developing in the large-amplitude cycle is on average 7.2 days, yielding a cycle period/delay ratio of ~ 3 . During the small-amplitude cycles, juvenile stage durations were estimated to be 42 days (lower 95% prediction limit 37 days; upper limit 48 days), which is significantly longer than the cycle period, producing a fractional period/delay ratio. Independent experiments that estimate the length-specific growth rates of juvenile *Daphnia* in populations at equilibrium (Supplementary Information) yield comparable values to those obtained from the individual bioassay experiment (Fig. 4b). Thus, two key predictions from theory were supported by results from these new experiments. First, coexisting attractors were observed in the same enrichment environment. Second, the stage durations of individual *Daphnia* were significantly shorter than the cycle period in the large-amplitude cycle, and significantly longer than the cycle period in the small-amplitude cycle (Fig. 4). To our knowledge, this is the first empirical support for cycles with a fractional period/delay ratio.

Theory on the dynamics of structured populations has provided new explanations for cycles and has enabled investigators to identify mechanisms giving rise to cycles of various types. Our experiments and measurements of individual performance in populations reveal a new type of small-amplitude cycle that arises from the interaction of instability in the predator-prey interaction causing fluctuations in resource density and a dynamically varying life-history feature of the predator. This interaction constrains the amplitude of the cycle even under nutrient-rich conditions where large-amplitude cycles prevail in either unstructured predator-prey models (for example, MacArthur-Rosenzweig models) or structured models with fixed stage durations for predators. The new cycles require the interaction of effects to emerge.

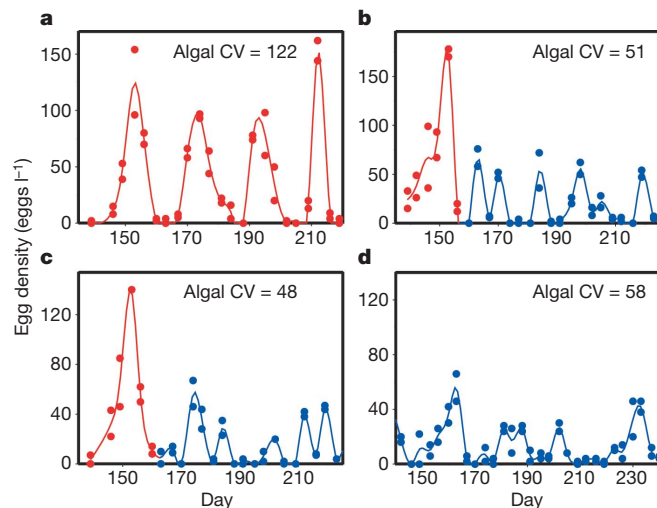


Figure 3 | Egg density dynamics during cycles. Dynamics of egg density in *Daphnia* populations displaying large-amplitude cycles (a) and small-amplitude cycles (b–d). The red line shows large-amplitude cycles as characterized by the fecundity probe (see Supplementary Information). The blue line shows small-amplitude cycles. The coefficient of variation (CV) in algal biomass is also presented.

Simple models containing nonlinearities in biological processes can produce phenomena such as alternative stable states or coexisting attractors, complex transients as dynamics move from the influence of one basin of attraction to another, and a range of cycles. A key question is whether these ‘rich’ results from nonlinear models actually provide fundamental insight into the dynamics of ecological

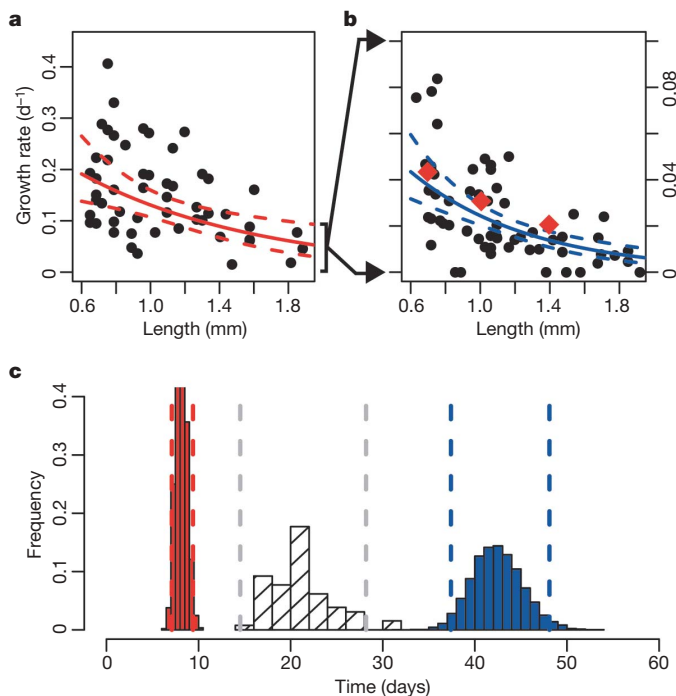


Figure 4 | Comparison of individual growth rates in large- and small-amplitude cycles. Length-specific growth rates in large-amplitude cycles (a) and small amplitude cycles (b) yield different estimates of juvenile stage duration (c; Supplementary Information). Diamonds show estimates from artificial cohort experiments (Supplementary Information). In large-amplitude cycles, juvenile stage duration (c, red) is significantly less than cycle periods (c, grey). In small-amplitude cycles, the juvenile duration is significantly greater than the cycle period (c, blue). Dashed lines show 95% prediction intervals (Supplementary Information).

systems²⁰. As we confront model predictions with well-studied experimental systems, where direct measurements of rates and manipulation of environmental conditions can be coupled quantitatively to models, more often than not, model expectations are upheld and general processes underlying the dynamics are being revealed. Structured population models of ecological interactions in particular are capturing the essence of dynamics of several 'model' systems (for example, *Tribolium*²¹, rotifers²², *Daphnia* and *Bosmina*¹², California redscale², Arctic char and brown trout²³). These models predict the occurrence of different dynamics as we move from one environment to the next, and identify key processes that lead to the different types of fluctuations. Although the models are costly to parameterize, the mechanisms apply to a broad range of organisms that share general life-history features, such as resource-dependent development rates and stage-specific interactions dictated by allometric scaling of energy acquisition and expenditure. Using these confrontations between theory and experiments, we are discovering important linkages between physiological ecology and life-history traits that explain dynamics of interacting populations in different environments.

METHODS SUMMARY

Daphnia-algal dynamics. Microcosms were established using techniques described previously⁴ and maintained in a constant environment at 24 °C with a 14:10 light cycle. Replicates were inoculated with a single clone of *Daphnia pulex* used in the previous experiments on food-dependent growth and reproduction²⁴. Initial nutrient enrichment with inorganic phosphorus and nitrogen placed the algal carrying capacity deep in parameter space where large-amplitude predator-prey cycles are predicted by theory⁴. Techniques for nutrient regeneration, non-destructive sampling of *Daphnia* populations, and chlorophyll analysis were carried out as described previously⁴. Single clones of each algal species from the University of Toronto Culture Collection were used. Biweekly samples were taken and no manipulations of *Daphnia* egg production were performed. An accurate estimate of periodicity in the population dynamics can be obtained by measuring the time period between reproduction bursts of fecund females in the population (Supplementary Information). Dynamics from 24 tanks were used to estimate this periodicity to compute the distribution shown in Fig. 4c.

Bioassay experiments. To estimate stage duration, we drew individuals of various sizes from *Daphnia* populations executing different types of cycles (large-amplitude and small-amplitude cycles) (Supplementary Information). One individual was placed in each flow-through vial, and each peristaltic pump could supply and remove water and food for ten replicate vials. A typical (1 mm) juvenile *Daphnia*, living at 24 °C, can filter at most 2 ml h⁻¹ (ref. 24). The flow rate through each vial was 40 ml h⁻¹, ensuring that the individual could not significantly alter the food concentration. Each individual experienced the same food concentration as its counterparts in the population of *Daphnia* in the microcosm. Individuals were drawn from populations during the period of egg production and followed for a minimum of 3 days, for individuals <1.4 mm, and 4 days for individuals >1.4 mm. An individual contributed only one estimate (that is, no repeated measures). Length-specific growth rates were calculated as $\ln(L_t/L_0)/\text{time}$ and regressed against length (mm) using nonlinear regression. Using these regression equations, the juvenile stage durations and 95% prediction intervals were estimated by determining the time required to grow from the size at birth (0.69 mm) to the size at first reproduction (1.6 mm) (Supplementary Information).

Received 21 May; accepted 27 June 2008.

1. Murdoch, W. W. Population regulation in theory and practise. *Ecology* **75**, 271–287 (1994).
2. Murdoch, W. W., Briggs, C. J. & Nisbet, R. M. *Consumer-Resource Dynamics* (Princeton Univ. Press, 2003).

3. de Roos, A. M. & Persson, L. Competition in size-structured populations, mechanisms inducing cohort formation and population cycles. *Theor. Popul. Biol.* **63**, 1–16 (2003).
4. McCauley, E., Nisbet, R. M., Murdoch, W. W., de Roos, A. M. & Gurney, W. S. C. Large amplitude cycles of *Daphnia* and its algal prey in enriched environments. *Nature* **402**, 653–656 (1999).
5. Bjornstad, O. N. & Grenfell, B. T. Noisy clockwork: Time-series analysis of population fluctuations in animals. *Science* **293**, 638–643 (2001).
6. McCauley, E. & Murdoch, W. W. Cyclic and stable populations: Plankton as paradigm. *Am. Nat.* **129**, 97–121 (1987).
7. McCauley, E. & Murdoch, W. W. Predator-prey dynamics in environments rich and poor in nutrients. *Nature* **343**, 455–457 (1990).
8. Claessen, D., de Roos, A. M. & Persson, L. Population dynamic theory of size-dependent cannibalism. *Proc. R. Soc. Lond. B* **271**, 333–340 (2004).
9. Medvinsky, A. B., Tikhonova, I. A., Li, B. L. & Malchow, H. Time delay as a key factor in model plankton dynamics. *C. R. Biologies* **327**, 277–282 (2004).
10. Hastings, A. & Wolkind, D. Age structure in predator-prey systems. *Theor. Pop. Biol.* **21**, 44–56 (1982).
11. De Roos, A. M., Metz, J. A. J., Evers, E. & Leipoldt, A. A size-dependent predator-prey interaction: who pursues whom? *J. Math. Biol.* **28**, 609–643 (1990).
12. McCauley, E., Nisbet, R. M., de Roos, A. M., Murdoch, W. W. & Gurney, W. S. C. Structured population models of herbivorous zooplankton. *Ecol. Monogr.* **66**, 479–501 (1996).
13. Nelson, W. A., McCauley, E. & Nisbet, R. M. Stage-structured cycles generate strong fitness-equalizing mechanisms. *Evol. Ecol.* **21**, 499–515 (2007).
14. Kretzschmar, M., Nisbet, R. M. & McCauley, E. A predator-prey model for zooplankton grazing on competing algal populations. *Theor. Pop. Biol.* **44**, 32–66 (1993).
15. Grover, J. P. Assembly rules for communities of nutrient-limited plants and specialist herbivores. *Am. Nat.* **143**, 258–282 (1994).
16. Abrams, P. A. & Holt, R. D. The impact of consumer-resource cycles on the coexistence of competing consumers. *Theor. Pop. Biol.* **62**, 281–295 (2002).
17. Murdoch, W. W. & McCauley, E. Three distinct types of dynamic behaviour shown by a single planktonic system. *Nature* **316**, 628–630 (1985).
18. Kendall, B. E. *et al.* Why do populations cycle? A synthesis of statistical and mechanistic modelling approaches. *Ecology* **80**, 1789–1805 (1999).
19. Murdoch, W. W. *et al.* Single-species models for many-species food webs. *Nature* **417**, 541–543 (2002).
20. Benton, T. G., Plaistow, S. J. & Coulson, T. N. Complex population dynamics and complex causation: devils, details and demography. *Proc. R. Soc. B* **273**, 1173–1181 (2006).
21. Dennis, B., Desharnais, R. A., Cushing, J. M., Henson, S. M. & Constantino, R. F. Estimating chaos and complex dynamics in an insect population. *Ecol. Monogr.* **71**, 277–303 (2001).
22. Yoshida, T., Jones, L. E., Ellner, S. P., Fussmann, G. F. & Hairston, N. G. Jr. Rapid evolution drives ecological dynamics in a predator-prey system. *Nature* **424**, 303–306 (2003).
23. Persson, L. *et al.* Culling prey promotes predator recovery-alternative states in a whole-lake experiment. *Science* **316**, 1743–1746 (2007).
24. Nisbet, R. M., McCauley, E., Gurney, W. S. C., Murdoch, W. W. & Wood, S. N. Formulating and testing a partially specified Dynamic Energy Budget Model. *Ecology* **85**, 3132–3139 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We acknowledge feedback from A. de Roos, J. Fox, J. Casas, L. Persson and C. Briggs. A. Potapov provided advice on the bifurcation analysis. Experiments and theoretical analysis were supported by NSERC (Discovery Grants and Accelerator Award), Canada Foundation for Innovation, and the Canada Research Chairs Program to E.M. W.A.N. acknowledges support from Alberta Ingenuity. R.M.N. acknowledges support from the US National Science Foundation (Grant DEB-0717259).

Author Contributions All authors contributed to the planning, execution and analysis of theory and experiments.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to E.M. (mccauley@ucalgary.ca).

LETTERS

GILT is a critical host factor for *Listeria monocytogenes* infection

Reshma Singh^{1,2}, Amanda Jamieson^{1†} & Peter Cresswell^{1,2}

Listeria monocytogenes is a Gram-positive, intracellular, food-borne pathogen that can cause severe illness in humans and animals. On infection, it is actively phagocytosed by macrophages¹; it then escapes from the phagosome, replicates in the cytosol, and subsequently spreads from cell to cell by a non-lytic mechanism driven by actin polymerization². Penetration of the phagosomal membrane is initiated by the secreted haemolysin listeriolysin O (LLO), which is essential for vacuolar escape *in vitro* and for virulence in animal models of infection³. Reduction is required to activate the lytic activity of LLO *in vitro*^{4–6}, and we show here that reduction by the enzyme γ -interferon-inducible lysosomal thiol reductase (GILT, also called Ifi30) is responsible for the activation of LLO *in vivo*. GILT is a soluble thiol reductase expressed constitutively within the lysosomes of antigen-presenting cells^{7,8}, and it accumulates in macrophage phagosomes as they mature into phagolysosomes⁹. The enzyme is delivered by a mannose-6-phosphate receptor-dependent mechanism to the endocytic pathway, where amino- and carboxy-terminal pro-peptides are cleaved to generate a 30-kDa mature enzyme^{7,8,10}. The active site of GILT contains two cysteine residues in a CXXC motif that catalyses the reduction of disulphide bonds^{7,8}. Mice lacking GILT are deficient in generating major histocompatibility complex class-II-restricted CD4⁺ T-cell responses to protein antigens that contain disulphide bonds^{11,12}. Here we show that these mice are resistant to *L. monocytogenes* infection. Replication of the organism in GILT-negative macrophages, or macrophages expressing an enzymatically inactive GILT mutant, is impaired because of delayed escape from the phagosome. GILT activates LLO within the phagosome by the thiol reductase mechanism shared by members of the thioredoxin family. In addition, purified GILT activates recombinant LLO, facilitating membrane permeabilization and red blood cell lysis. The data show that GILT is a critical host factor that facilitates *L. monocytogenes* infection.

GILT is the only known thiol oxidoreductase present in phagosomes and we speculated that it might activate LLO *in vivo*. Consistent with this idea, when we infected wild type and GILT-deficient mice with *L. monocytogenes* we observed more rapid bacterial clearance from the spleen and liver of the GILT-deficient mice (Fig. 1a). Initial infection was as effective as in wild-type mice on the basis of the similar bacterial colony-forming units (c.f.u.) observed at day one after infection. No difference in the clearance rates of *Salmonella typhimurium* was seen between wild type and GILT knockout mice (data not shown). To determine if the difference in clearance was a function of defective intracellular growth, we examined bacterial growth in bone-marrow-derived macrophages *in vitro*. Replication of *L. monocytogenes* was clearly impaired in macrophages from GILT-deficient mice (Fig. 1b). This was not due to an inherent defect in phagocytosis or bacterial killing by GILT^{−/−}

macrophages, which cleared an *in vitro* infection by non-pathogenic *Escherichia coli* as efficiently as wild-type macrophages (Supplementary Fig. 1). Nor was it due to a problem in phagosomal acidification—acidification of both lysosomes and phagosomes occurred with the same kinetics in GILT^{−/−} and wild-type macrophages (Supplementary Fig. 2).

Because GILT is active in phagosomes it seemed likely that the growth deficiency was due to a defect in phagosomal escape. Once in the cytosol, *L. monocytogenes* polymerizes host actin, which allows movement through the cytosol and facilitates intercellular spread²,

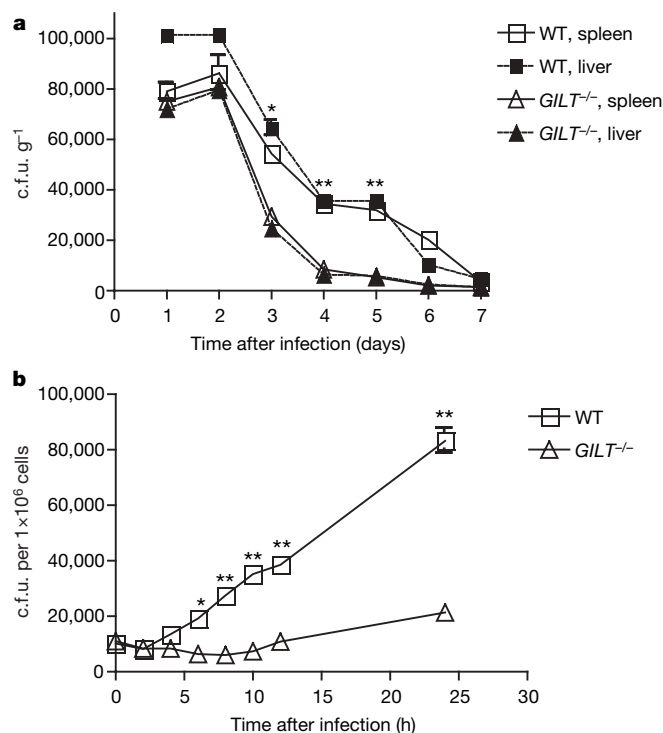


Figure 1 | Growth of *L. monocytogenes* is decreased in GILT-deficient mice and GILT-deficient macrophages. **a**, *In vivo* colony counts of *L. monocytogenes* (strain 10403s) injected intravenously in wild-type (WT) C57BL/6 or GILT-deficient mice. At each time point three mice per group were killed and spleens and livers were harvested each day for bacterial colony counts to obtain c.f.u. g^{−1}. A representative of three individual experiments is shown. **b**, *In vitro* growth of *L. monocytogenes* in bone-marrow-derived macrophages from wild-type C57BL/6 and GILT-deficient mice. The cells (10⁶ per well) were lysed at each time point shown and plated to obtain the c.f.u. A representative of three individual experiments is shown. Asterisk, $P < 0.05$; double asterisk, $P < 0.01$. Error bars in **a** and **b** indicate mean \pm standard error.

¹Department of Immunobiology, ²Howard Hughes Medical Institute, Yale University School of Medicine, 300 Cedar Street, New Haven, Connecticut 06250-8011, USA. †Present address: Max F. Perutz Laboratories GmbH, Rennweg 95a, 1030 Wien 3, Landstraße, Wiene A1030, Austria.

and actin polymerization provides a convenient assay for cytosolic entry. Indeed, we saw a delay in actin polymerization, detected by phalloidin staining, in infected GILT-deficient macrophages, consistent with impaired escape from the phagosome (Fig. 2a, b). In wild-type macrophages actin polymerization was observed as early as 30 min after infection. This typically results in overwhelming infection and cell death within 36 to 48 h. In GILT-deficient macrophages, actin polymerization was not observed until 6–8 h. When wild type or GILT-deficient macrophages were infected with an LLO-negative *L. monocytogenes* strain the rate of actin polymerization was reduced to a level below that seen with wild-type bacteria in the GILT-negative cells (Fig. 2a, b). Consistent with a role for GILT in *L. monocytogenes* infection, a marked increase in cytosolic access was observed on infection of the GILT-negative human promonocytic cell line THP-1 when human GILT was expressed by retroviral transduction¹³ (Supplementary Fig. 3). GILT is induced in THP-1 cells on exposure to bacteria, but more than 24 h is required¹³.

To confirm the escape defect we examined infected cells by electron microscopy (Fig. 2c). After 2 h virtually no bacteria were seen outside of membrane-bound phagosomes in GILT-deficient macrophages, and only limited actin polymerization was observed even 8 h

after infection. Furthermore, defective phagosomal escape, detected by the delay in the induction of actin polymerization, was reversed when GILT was expressed in GILT-negative macrophages by retroviral transduction (Fig. 2d, e). However, expression of a double cysteine mutant of GILT (C69S/C71S), lacking both active site cysteine residues and unable to catalyse disulphide bond reduction¹¹, failed to reverse the escape defect.

LLO has a single cysteine residue at position 485 and a mutant with an alanine substitution at that position does not require reduction for activation¹⁴. To determine if the mutation also reversed GILT dependence, we repeated the *in vitro* infection experiments described above with an *L. monocytogenes* strain expressing this LLO variant. For this organism growth was equivalent in wild type and *GILT*^{-/-} macrophages (Fig. 3a), as was the rate of escape from the phagosome (Fig. 3b, c). We also examined the capacity of recombinant enzymatically active precursor GILT^{8,10,15} to induce lytic activity in purified recombinant LLO (characterized in Supplementary Fig. 4). GILT clearly was able to activate LLO, determined by lysis of purified bone-marrow-derived macrophages (Fig. 3d) and haemolysis of sheep red blood cells (Fig. 3e). Similar lytic curves for macrophages were obtained using trypan blue exclusion (data not shown). Lysis

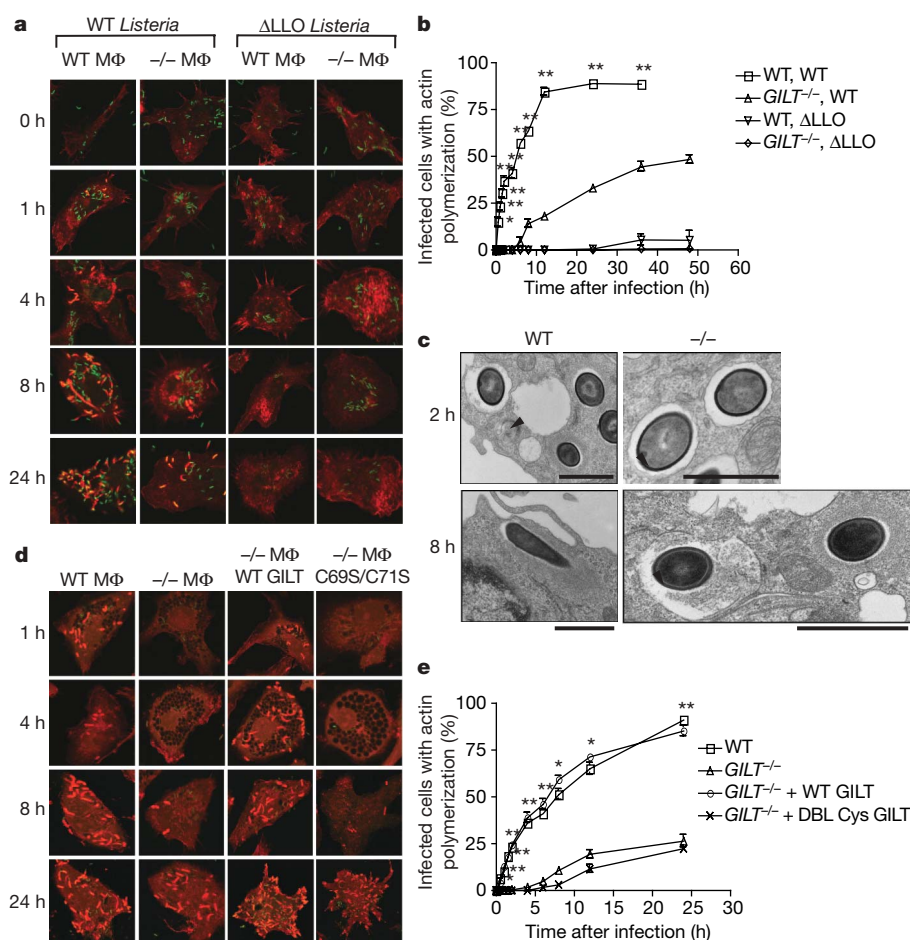


Figure 2 | Phagosomal escape of *L. monocytogenes* is delayed in GILT-negative macrophages. **a**, Actin polymerization in wild type and *GILT* knockout macrophages (MΦ) infected with wild-type *L. monocytogenes* or ΔLLO *L. monocytogenes*. Both strains of bacteria express GFP and are green. Phalloidin staining of actin is in red. **b**, Quantification of actin polymerization. A total of 1,000 cells per sample was counted for each time point. The data derive from the average of three experiments and statistical significance was assessed by comparing the infection of wild type and *GILT*^{-/-} macrophages with wild-type bacteria. Asterisk, $P < 0.05$; double asterisk, $P < 0.01$. Error bars indicate mean \pm standard error. **c**, Transmission electron micrographs of wild type or *GILT*^{-/-}-infected macrophages. The arrowheads indicate intact phagosomal membrane, and

the reordering of the cytoplasm in the initial stages of actin polymerization can be seen around bacteria in the cytosol. Scale bar, 1 μ m. No arrowheads are present in the 8 h image for wild-type cells because no bacteria are contained within membrane-bound compartments. **d**, Actin polymerization in infected wild type and *GILT*-negative macrophages, and *GILT*-negative macrophages expressing GILT or inactive mutant GILT (C69S/C71S). **e**, Quantification of actin polymerization in macrophages infected as in **d**. The data derive from the average of three experiments and statistical significance was assessed by comparing the wild type and untransduced *GILT*^{-/-} macrophages and, separately, the *GILT*^{-/-} macrophages transduced with wild type and mutant GILT retroviruses. Asterisk, $P < 0.05$; double asterisk, $P < 0.01$. Error bars indicate mean \pm standard error.

was abrogated if the enzymatic activity of GILT was eliminated by pre-treatment with the thiol-reactive reagent *N*-ethylmaleimide (NEM).

GILT shares with thioredoxin a reduction mechanism in which the N-terminal cysteine residue in the CXXC active site reduces a substrate disulphide bond by a nucleophilic attack on one of the involved cysteine residues. This generates a disulphide-linked mixed enzyme-substrate intermediate that is rapidly resolved by an attack of the second active site cysteine residue on the first^{7,8}. The second step in the reaction, known as the escape pathway¹⁶, can be prevented by mutation of the second active site cysteine, generating a 'trapping mutant' that allows mixed enzyme-substrate dimers to be isolated. To determine whether GILT uses this mechanism to activate LLO, bone-marrow-derived macrophages from wild type and *GILT*^{-/-} mice were again infected with *L. monocytogenes*, together with *GILT*^{-/-} macrophages that were retrovirally transduced with a GILT trapping mutant (C71S). Infected cells were lysed in detergent and immunoprecipitated GILT was subjected to SDS-polyacrylamide gel electrophoresis (PAGE) followed by western blotting to detect associated LLO. LLO co-precipitated with GILT only from infected macrophages expressing the trapping mutant (Fig. 4a). Immunofluorescence microscopy confirmed that in wild-type macrophages the organism co-localizes with GILT in LAMP-1-positive phagosomes (data not shown). Thus, GILT uses the classical thiol reductase mechanism to activate LLO in phagosomes and initiate the escape of *L. monocytogenes* to the cytosol.

In the absence of LLO, *L. monocytogenes* may eventually escape from the phagosome, just as in the absence of GILT (Fig. 2a). However, when active GILT is present escape of LLO-positive

bacteria is extremely rapid. The identity of the disulphide bond targeted by GILT is unclear. The single cysteine present in LLO lies in a short tryptophan-rich sequence that initiates pore formation on binding to membrane cholesterol. It has been suggested that a small thiol-containing molecule is disulphide linked to the cysteine residue, and that this inhibits activation of the haemolysins¹⁷. The equivalent cysteine residue in perfringolysin O (PFO), a related haemolysin derived from *Clostridium perfringens*, is not necessary for binding to cholesterol in the membrane, but is necessary for the formation of a pre-pore complex^{18–20}. GILT-mediated exposure of the critical cysteine residue in LLO may facilitate a conformational change that allows the formation of the pre-pore complex and full activation. Notably, the ability of GILT to activate haemolysins is not limited to LLO: GILT can also activate streptolysin O (SLO), derived from *Streptococcus pyogenes*, as measured by the haemolysis of sheep red blood cells, but it is not required to activate an SLO mutant that lacks the characteristic single cysteine residue²¹ (Supplementary Fig. 5).

In most cell types GILT can be induced by interferon (IFN)- γ , and IFN- γ induction during the early stages of infection may enhance the ability of *L. monocytogenes* to infect other cells, including hepatocytes where it also replicates *in vivo*^{22,23}. The ability of GILT to activate SLO as well as LLO suggests that it may activate other members of this highly conserved family of haemolysins. Phagocytosis of haemolysin-expressing organisms may not be essential for activation of the lytic activity, as we have recently found that Toll-like receptor (TLR)-mediated activation of macrophages by *E. coli* lipopolysaccharide induces secretion of the enzymatically active precursor form of GILT^{13,24} that activates SLO *in vitro* (Supplementary Fig. 5). GILT is functional even at neutral pH, retaining approximately 30% of the

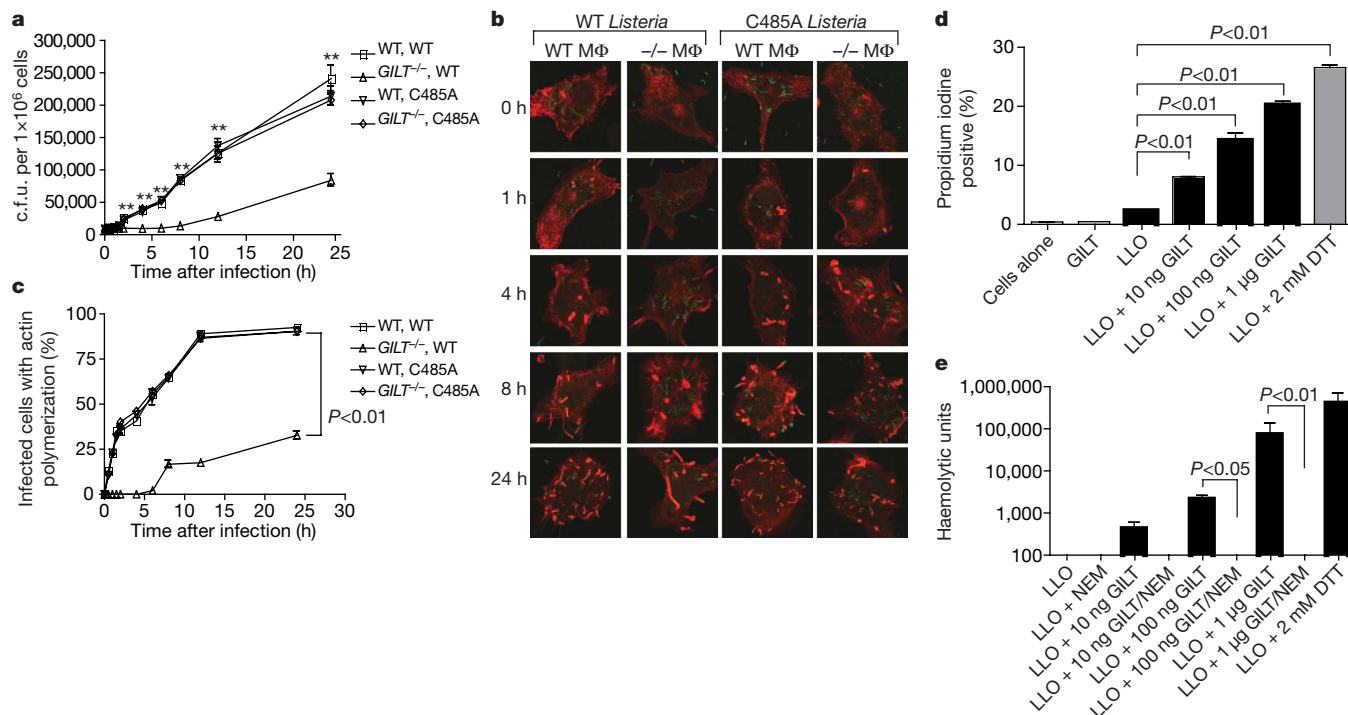


Figure 3 | LLO is activated by GILT reduction during infection and in cell-free assays. **a**, *In vitro* growth of wild type and C485A LLO mutant *L. monocytogenes* in wild type and *GILT*^{-/-} macrophages. The experiment, performed as described in Fig. 1b, was repeated three times and a representative experiment is shown. Asterisk, *P* < 0.05; double asterisk, *P* < 0.01. **b**, Actin polymerization in wild type and *GILT*^{-/-} macrophages infected with wild-type *L. monocytogenes* or C485A LLO mutant *L. monocytogenes*. The bacteria were detected by immunofluorescence (green) and phalloidin staining of actin is in red. **c**, Quantification of actin polymerization as described in Fig. 2b. The data represent an average of three independent experiments; all points after 0.5 h are statistically significant to

P < 0.01. **d**, Viability of GILT-negative macrophages assessed by propidium iodide staining after incubation of 1 × 10⁵ cells with 300 ng LLO for 30–45 min. Cells were incubated alone, with activated GILT (1 μg) and 25 μM dithiothreitol (DTT), with unactivated LLO and 25 μM DTT, with DTT-activated LLO, or with LLO and 1 μg, 100 ng or 10 ng activated GILT. The data represent the average of three independent experiments and *P* values for the samples with significant differences are shown. **e**, Haemolytic activity of purified LLO pre-incubated with active GILT or GILT inactivated by NEM treatment. The data represent the average of three independent experiments and *P* values for the samples with significant differences are shown. All error bars indicate mean ± standard error.

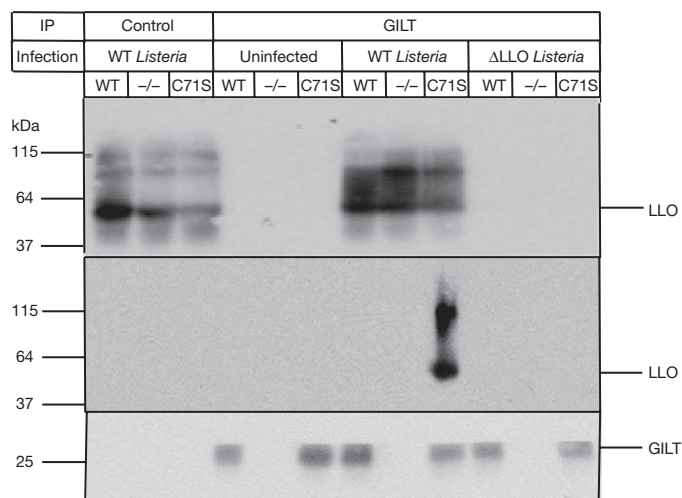


Figure 4 | GILT activates LLO by using the classical thioredoxin reduction mechanism. Wild-type macrophages, *GILT*^{-/-} macrophages, or *GILT*^{-/-} macrophages reconstituted with the C71S GILT trapping mutant were uninfected or infected with either wild-type *L. monocytogenes* or ΔLLO *L. monocytogenes* for 2 h before detergent solubilization, immunoprecipitation (IP) and SDS-PAGE. Top panel: western blot of LLO in the detergent extracts. Middle panel: western blot with an LLO-specific antiserum of immunoprecipitates isolated from the extracts using a control monoclonal antibody (left three lanes) or an anti-GILT monoclonal antibody (right nine lanes). LLO is only co-precipitated with the C71S trapping mutant. Bottom panel: western blot for GILT in the immunoprecipitates.

activity seen at pH 4.5 (refs 7 and 8) and it has been reported that the local pH can be as low as 5.7 at sites of bacterial infection²⁵. Secretion of GILT by macrophages at such a site could facilitate local haemolysis-mediated tissue damage, perhaps including lysis of inflammatory cells recruited for the purpose of host defence.

METHODS SUMMARY

The *L. monocytogenes* strains used are listed in the Methods. Infections in C57BL/6 and C57BL/6 *GILT* knockout mice were performed as described²⁶. Bone-marrow-derived macrophages were prepared as described²⁴. *In vitro* infections were done in the presence of gentamicin to prevent extracellular bacterial growth, and at a multiplicity of infection (MOI) of 5 (unless otherwise stated in the Methods).

Electron microscopy and immunofluorescence experiments and antibodies used are as described in the Methods. Co-localization and quantification were performed by direct visualization on a Leica DMIRE2 confocal microscope. Figure assembly was done with Adobe Photoshop and Adobe Illustrator.

LLO protein was purified according to published protocols with modifications described in the Methods²⁷. The mean ± standard error is shown in the figures and *P* values were calculated using a two-tailed two-sample equal variance Student's *t*-test. A *P* value of less than 0.05 was determined to be statistically significant.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 12 March; accepted 12 August 2008.
Published online 24 September 2008.

- Pizarro-Cerda, J., Sousa, S. & Cossart, P. Exploitation of host cell cytoskeleton and signalling during *Listeria monocytogenes* entry into mammalian cells. *C. R. Biol.* **327**, 523–531 (2004).
- Tilney, L. G. & Portnoy, D. A. Actin filaments and the growth, movement, and spread of the intracellular bacterial parasite, *Listeria monocytogenes*. *J. Cell Biol.* **109**, 1597–1608 (1989).
- Barry, R. A., Bouwer, H. G., Portnoy, D. A. & Hinrichs, D. J. Pathogenicity and immunogenicity of *Listeria monocytogenes* small-plaque mutants defective for intracellular growth and cell-to-cell spread. *Infect. Immun.* **60**, 1625–1632 (1992).
- Billington, S. J., Jost, B. H. & Songer, J. G. Thiol-activated cytolysins: structure, function and role in pathogenesis. *FEMS Microbiol. Lett.* **182**, 197–205 (2000).

- Geoffroy, C., Gaillard, J. L., Alouf, J. E. & Berche, P. Purification, characterization, and toxicity of the sulfhydryl-activated hemolysin listeriolysin O from *Listeria monocytogenes*. *Infect. Immun.* **55**, 1641–1646 (1987).
- Portnoy, D. A., Chakraborty, T., Goebel, W. & Cossart, P. Molecular determinants of *Listeria monocytogenes* pathogenesis. *Infect. Immun.* **60**, 1263–1267 (1992).
- Arunachalam, B., Phan, U. T., Geuze, H. J. & Cresswell, P. Enzymatic reduction of disulfide bonds in lysosomes: characterization of a gamma-interferon-inducible lysosomal thiol reductase (GILT). *Proc. Natl Acad. Sci. USA* **97**, 745–750 (2000).
- Phan, U. T., Arunachalam, B. & Cresswell, P. Gamma-interferon-inducible lysosomal thiol reductase (GILT). Maturation, activity, and mechanism of action. *J. Biol. Chem.* **275**, 25907–25914 (2000).
- Garin, J. *et al.* The phagosome proteome: insight into phagosome functions. *J. Cell Biol.* **152**, 165–180 (2001).
- Phan, U. T., Lackman, R. L. & Cresswell, P. Role of the C-terminal propeptide in the activity and maturation of gamma-interferon-inducible lysosomal thiol reductase (GILT). *Proc. Natl Acad. Sci. USA* **99**, 12298–12303 (2002).
- Maric, M. *et al.* Defective antigen processing in GILT-free mice. *Science* **294**, 1361–1365 (2001).
- Sealy, R. *et al.* Target peptide sequence within infectious human immunodeficiency virus type 1 does not ensure envelope-specific T-helper cell reactivation: influences of cysteine protease and gamma interferon-induced thiol reductase activities. *Clin. Vaccine Immunol.* **15**, 713–719 (2008).
- Lackman, R. L. & Cresswell, P. Exposure of the promonocytic cell line THP-1 to *Escherichia coli* induces IFN-γ-inducible lysosomal thiol reductase expression by inflammatory cytokines. *J. Immunol.* **177**, 4833–4840 (2006).
- Michel, E., Reich, K. A., Favier, R., Berche, P. & Cossart, P. Attenuated mutants of the intracellular bacterium *Listeria monocytogenes* obtained by single amino acid substitutions in listeriolysin O. *Mol. Microbiol.* **4**, 2167–2178 (1990).
- Hastings, K. T., Lackman, R. L. & Cresswell, P. Functional requirements for the lysosomal thiol reductase GILT in MHC class II-restricted antigen processing. *J. Immunol.* **177**, 8569–8577 (2006).
- Walker, K. W. & Gilbert, H. F. Scanning and escape during protein-disulfide isomerase-assisted protein folding. *J. Biol. Chem.* **272**, 8845–8848 (1997).
- Alouf, J. E., Billington, S. J. & Jost, B. H. in *Bacterial Toxins: A Comprehensive Sourcebook* 643–658 (Academic, 2005).
- Heuck, A. P., Tweten, R. K. & Johnson, A. E. Assembly and topography of the prepore complex in cholesterol-dependent cytolysins. *J. Biol. Chem.* **278**, 31218–31225 (2003).
- Soltani, C. E., Hotze, E. M., Johnson, A. E. & Tweten, R. K. Structural elements of the cholesterol-dependent cytolysins that are responsible for their cholesterol-sensitive membrane interactions. *Proc. Natl Acad. Sci. USA* **104**, 20226–20231 (2007).
- Soltani, C. E., Hotze, E. M., Johnson, A. E. & Tweten, R. K. Specific protein-membrane contacts are required for prepore and pore assembly by a cholesterol-dependent cytolysin. *J. Biol. Chem.* **282**, 15709–15716 (2007).
- Pinkney, M., Beachey, E. & Kehoe, M. The thiol-activated toxin streptolysin O does not require a thiol group for cytolytic activity. *Infect. Immun.* **57**, 2553–2558 (1989).
- Bouwer, H. G. *et al.* *Listeria monocytogenes*-infected hepatocytes are targets of major histocompatibility complex class Ib-restricted antilisterial cytotoxic T lymphocytes. *Infect. Immun.* **66**, 2814–2817 (1998).
- Haschtmann, D., Gerber, H. J. & Mielke, M. E. Cytotoxic activity of murine resident peritoneal cells against *Listeria monocytogenes*-infected hepatocytes *in vitro*. *Microbes Infect.* **7**, 1177–1183 (2005).
- Lackman, R. L., Jamieson, A. M., Griffith, J. M., Geuze, H. & Cresswell, P. Innate immune recognition triggers secretion of lysosomal enzymes by macrophages. *Traffic* **8**, 1179–1189 (2007).
- Bryant, R. E., Rashad, A. L., Mazza, J. A. & Hammond, D. beta-Lactamase activity in human pus. *J. Infect. Dis.* **142**, 594–601 (1980).
- Pamer, E. G., Wang, C. R., Flaherty, L., Lindahl, K. F. & Bevan, M. J. H-2M3 presents a *Listeria monocytogenes* peptide to cytotoxic T lymphocytes. *Cell* **70**, 215–223 (1992).
- Gedde, M. M., Higgins, D. E., Tilney, L. G. & Portnoy, D. A. Role of listeriolysin O in cell-to-cell spread of *Listeria monocytogenes*. *Infect. Immun.* **68**, 999–1003 (2000).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful to D. Portnoy for advice and reagents, and to N. Dometios for manuscript preparation. We acknowledge the valuable contribution of the late M. Pypaert to the electron microscopy. This work was supported by NIH AI023081 (P.C.) and the Howard Hughes Medical Institute (P.C., R.S.).

Author Contributions R.S. performed experiments, A.J. assisted with the *in vivo* infection, and R.S. and P.C. wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to P.C. (peter.cresswell@yale.edu).

METHODS

Bacterial strains. *Listeria monocytogenes* was grown in brain heart infusion (BHI) broth (BD). The strains used were: wild type (10403s), *L. monocytogenes*-GFP (1039), *L. monocytogenes* Δ LLO (1039 Δ LLO) (gifts from H. Agaisse)²⁸ and DP-L4391 (C485A LLO) (a gift from D. Portnoy). LLO was purified from the DP-E3570 *E. coli* strain (donated by D. Portnoy) grown in Luria-Bertani (LB) broth (BD) supplemented with kanamycin. Production of LLO was induced by 1 mM isopropyl- β -D-thiogalactopyranoside (IPTG) at 30 °C. *E. coli* used for infection was grown in LB supplemented with ampicillin. **In vivo bacterial infections.** Infections of C57BL/6 and C57BL/6 GILT-deficient mice were performed as described previously²⁶. Six-eight-week-old mice were used for each experiment.

Cell culture. Bone-marrow-derived macrophages were cultured as described²⁴. Briefly, bone marrow was harvested from the femurs of 8–10-week-old mice and cells cultured for 5–6 days in RPMI 1640 containing 20% fetal calf serum (Hyclone), 100 units ml⁻¹ penicillin (GIBCO), 100 μ g ml⁻¹ streptomycin (GIBCO), 10 mM HEPES (GIBCO), 1% non-essential amino acids (GIBCO), 2 mM L-glutamine (GIBCO), 1 mM sodium pyruvate, and 0.035% β -mercaptoethanol, supplemented with 10 ng ml⁻¹ granulocyte-monocyte colony stimulating factor. HEK 293 cells were grown in DMEM with 10% bovine calf serum (Hyclone).

In vitro infections. Infections of macrophages for *in vitro* growth assays were at an MOI of 5 except for Fig. 1b where an MOI of 0.1 was used and Fig. 3a when an MOI of 0.5 was used. For infection, 16 h *L. monocytogenes* cultures were diluted 1:10 in fresh BHI and grown for an additional 2 h at 37 °C with shaking. Aliquots of mid-log phase bacteria ($\sim 5 \times 10^8$ c.f.u. ml⁻¹) were washed once in PBS and used to infect macrophages in medium without antibiotics. After 30 min at 37 °C gentamicin (50 μ g ml⁻¹) was added and the cells incubated for another 30 min at 37 °C, washed, and incubated in 1 ml fresh medium at 37 °C. Cells were lysed at 0, 2, 4, 6, 8, 10, 12 and 24 h in 1 ml water for 5–10 min. Serial dilutions were plated on BHI plates containing chloramphenicol and colonies were counted the next day to determine c.f.u.

Transmission electron microscopy and immunofluorescence. For transmission electron microscopy, cells were fixed in 0.2% glutaraldehyde in 0.1 M cacodylate buffer for 1 h at room temperature and processed as described²⁴. For immunofluorescence, macrophages were plated on coverslips 12 h before infection. The cells were fixed in 2% glutaraldehyde for 20 min at room temperature at each time point. For phalloidin staining, the coverslips were washed in PBS and permeabilized with 0.1% Triton X-100 in PBS for 3 min. After washing, methanolic phalloidin stock solution was diluted 40 \times into 1% BSA in PBS and the coverslips stained for 20 min at room temperature. Intracellular antibody staining was done after cell permeabilization in 0.1% saponin for 20 min at room temperature. Primary antibodies were added for 30 min and secondary antibodies for a further 30 min. The antibodies used were: R.mGILT¹¹, MaP.mGILT6 (mouse anti-mouse GILT monoclonal antibody)²⁴, phalloidin conjugated to Alexa 546 (Molecular Probes), and rabbit anti-*Listeria* sp. FITC (Affinity BioReagents). Alexa Fluor 546 and 633 conjugated secondary antibodies were used (Molecular Probes). Coverslips were mounted using ProLong Gold mounting solution (Molecular Probes).

Retroviral constructs and spinfection. Wild-type mouse GILT cDNA was cloned into the pLPCX vector (Clontech) using BglII and HindIII restriction enzymes. GILT mutants were made by site-directed mutagenesis with the following primers: C69S/C71S, 5'-GAGTCCCTGTCCGGAGCTAGCCGCTACTT

CCTCCG-3', and C71S, 5'-CCCTGTGTGGAGCTAGCCGCTACTTCCTC-3'. To produce retrovirus, HEK 293 cells were transfected with 12 μ g of each pLPCX construct and 12 μ g pCL-Eco with 60 μ l Lipofectamine 2000 (Invitrogen). After 12 h the medium was changed to macrophage culture medium and the cells were shifted to 32 °C. After 24 h filtered supernatant was added to day 2 macrophage cultures. The cells were spininfected at 32 °C, at 2,900 r.p.m. for 90 min and cultured at 37 °C and differentiated as usual.

LLO purification. Purification of LLO was as described except that the β -mercaptoethanol was omitted from all the buffers and 0.5 mM DTT was used for storage²⁷.

Cell lytic assays. Precursor GILT, purified from supernatants of baculovirus-infected insect cells as described¹¹, was activated with 25 μ M DTT at room temperature for 10 min, and for some experiments inactivated with NEM (1.5 mM) followed by dialysis against normal saline (pH 5.5). It was then incubated with LLO at 37 °C for 30 min. An aliquot of LLO was activated using DTT (2 mM) as a positive control. The samples were incubated with macrophages at 37 °C for 30 min. Viability was assessed by propidium iodide staining and FACS analysis. For haemolysis assays LLO was added to sheep red blood cells (Innovative Research) in PBS, pH 5.5, in 96-well plates on ice. After 30 min at 37 °C lysis was determined spectrophotometrically, with Triton X-100 lysis serving as 100% release. One haemolytic unit (HU) is the amount of toxin that releases half the haemoglobin. SLO (Aalto Bio Reagents Ltd.) was activated by 4 mM DTT. Purified SLO with a cysteine to alanine mutation at position 530 (ref. 21) was a gift from N. Andrews.

LLO co-immunoprecipitation. Macrophages were infected with *L. monocytogenes* for 2 h and extracted in Tris-buffered saline, pH 7.4, containing 1% Triton X-100, protease inhibitors (Roche), and 10 mM methyl methanethiolsulphonate for 30 min on ice. GILT was immunoprecipitated using the monoclonal antibody MaP.mGILT6, the samples separated by reducing SDS-PAGE and LLO detected by western blot using a rabbit anti-LLO antibody and goat anti-rabbit immunoglobulin conjugated with horseradish peroxidase (HRP). An anti-H2-K^b monoclonal antibody, Y-3, was used as a control. The blot was developed with ECL reagents.

Lysosomal and phagosomal acidification. Lysosomal pH was determined after uptake of Oregon-green-labelled dextran (10,000 Da molecular mass, 25 μ g ml⁻¹; Molecular Probes) over 45 min on ice as described^{29,30}. Cells were extensively washed in PBS at neutral pH and read on a SpectraMax M5 plate reader (Molecular Devices), and analysed with SoftmaxPro software (Molecular Devices). Fluorescent emission at 520 nm was measured with alternating excitation at 450 nm and 490 nm for 40 min. Conversion of the 450/490 excitation ratio to pH was calculated based on a standard curve generated using excitation ratios of the dextran in standard buffers. Phagosomal pH was determined after the uptake of 3 μ m carboxy beads (Polysciences Inc.) covalently labelled with carboxyfluorescein-SE (Molecular Probes) using a similar approach.

28. Agaisse, H. *et al.* Genome-wide RNAi screen for host factors required for intracellular bacterial infection. *Science* **309**, 1248–1251 (2005).
29. Yates, R. M., Hermetter, A. & Russell, D. G. The kinetics of phagosome maturation as a function of phagosome/lysosome fusion and acquisition of hydrolytic activity. *Traffic* **6**, 413–420 (2005).
30. Yates, R. M., Hermetter, A., Taylor, G. A. & Russell, D. G. Macrophage activation downregulates the degradative capacity of the phagosome. *Traffic* **8**, 241–250 (2007).

LETTERS

Non-random segregation of sister chromosomes in *Escherichia coli*

Martin A. White¹, John K. Eykelenboom¹, Manuel A. Lopez-Vernaza¹, Emily Wilson¹ & David R. F. Leach¹

It has long been known that the 5' to 3' polarity of DNA synthesis results in both a leading and lagging strand at all replication forks¹. Until now, however, there has been no evidence that leading or lagging strands are spatially organized in any way within a cell. Here we show that chromosome segregation in *Escherichia coli* is not random but is driven in a manner that results in the leading and lagging strands being addressed to particular cellular destinations. These destinations are consistent with the known patterns of chromosome segregation^{2,3}. Our work demonstrates a new level of organization relating to the replication and segregation of the *E. coli* chromosome.

Prokaryotic cells were long considered to be featureless until recent advances in imaging revealed an array of internal structures and sub-cellular organizations⁴. One such example is the nuclear architecture of *E. coli*, where the domains of the left and right chromosome arms occupy distinct cellular locations². During replication, these domains are progressively segregated by an unknown mechanism that results in translational symmetry of the chromosome arms (known as replichores)^{2,3}. Counter-intuitively, to obtain this translational symmetry there must be mirror symmetry in the segregation of the leading and lagging strands of the two replication forks (Fig. 1A). This leads to three possible situations. Either the two lagging strands of replication are positioned at mid-cell while the leading strands migrate to the cell poles; the two lagging strands migrate to the cell poles while the leading strands are positioned at mid-cell; or a random combination of these two segregation patterns occurs within the population. The

specific hypothesis that leading strands segregate to cell poles has been proposed^{5,6} to explain the preponderance of highly expressed genes on the leading strands of the left (L) and right (R) replichores⁷. However, no evidence for or against this has yet been presented.

To distinguish these three possibilities, we designed a construct (Supplementary Fig. 1) that would allow us specifically to visualize a locus on the right replichore (*lacZ*) that was replicated on the leading strand of the replication fork (R_{lead}). This is accomplished by inducing SbcCD-mediated palindrome cleavage⁸ in a *recA*⁻ mutant that has the palindrome flanked upstream by an array of *tetO* sites and downstream by an array of *lacO* sites. These arrays are visualized by fluorescence microscopy upon the binding of TetR–yellow fluorescent protein (YFP) and LacI–cyan fluorescent protein (CFP), respectively, and allow us to follow the cellular position of the chromosomal region containing the palindrome as it is replicated and segregated. This is visible as a twin YFP–CFP spot, demonstrating the presence of DNA on both sides of the palindrome. Induction of SbcCD results in the specific cleavage of the palindrome that was replicated on the lagging strand⁸ (R_{lag}) by formation of a DNA hairpin⁹. In a *recA*⁻ mutant the broken chromosome is degraded¹⁰ leaving behind the intact copy of the construct located on R_{lead} (Fig. 1B). As long as the labelling method does not disrupt the known pattern of chromosome segregation, then the location of only one of the chromosome arms needs to be known for the rest to be inferred (Fig. 1A).

The cellular position of this construct was followed during growth, and the number of segregated copies related to cell length. In the absence of induced double-strand breaks, 68.8% of exponentially growing cells longer than 1.50 μm had two visibly segregated copies of the construct. In these cells, the loci were segregated such that one was located at mid-cell and the other at a cell pole (Fig. 2a). This

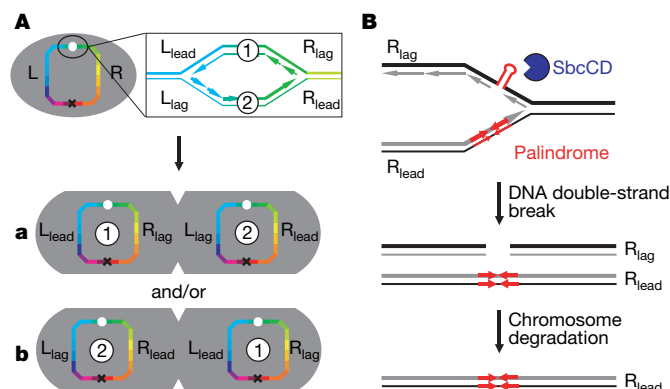


Figure 1 | Distinguishing leading and lagging strands. **A**, Cartoon demonstrating how mirror symmetry of leading and lagging strands accounts for translational symmetry of sister chromosomes. **Aa**, Segregation of leading strands towards the cell poles and of lagging strands towards mid-cell. **Ab**, Segregation of lagging strands towards the cell poles and of leading strands towards mid-cell. **B**, SbcCD cleaves a DNA hairpin formed by the palindrome on the lagging strand of replication, and the broken chromosome is degraded in a *recA*⁻ mutant.

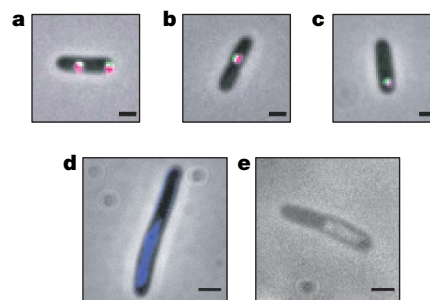


Figure 2 | Visualization of construct. *E. coli* cell containing: **a**, two segregated copies of the construct; **b**, one located at mid-cell; and **c**, one located at a cell pole. The CFP signal is pseudocoloured green, YFP magenta. Overlapping CFP and YFP signals appear white. Visualization of the nucleoid in cells subjected to SbcCD-mediated palindrome cleavage for 120 min by DAPI staining (DNA, blue) (**d**) and mounting on gelatin (DNA, white) (**e**). Scale bars, 1 μm (**a–c**) and 2 μm (**d, e**).

¹Institute of Cell Biology, School of Biological Sciences, University of Edinburgh, King's Buildings, Edinburgh EH9 3JR, UK.

localization pattern is consistent with published data on the pattern of chromosome segregation and localization in *E. coli*²⁻³ and was indistinguishable from the *recA*⁺ parental strains. Localization was also not independently affected by either induction of SbcCD or the presence of the palindrome. The induction of SbcCD expression in a palindrome-containing strain, however, caused a decrease in the number of cells longer than 1.50 μm with two visibly segregated copies of the construct over time (Fig. 3a). This decrease was strongly associated with an increase in the number of cells with only one visible copy of the construct, suggesting that cleavage and degradation of one of the replicated chromosomes had occurred. This was not observed in the control strain lacking the palindrome whereby 80.2% \pm 3.1 (range, $n = 574$) of cells had two visibly segregated copies of the construct at T_0 and 81.2% \pm 1.8 (range, $n = 796$) after 90 min of SbcCD induction. Interestingly, visualization of the nucleoid by either DAPI staining or mounting cells on gelatin¹¹ revealed that even after extensive cleavage and degradation, the intact chromosome remains in one cell half (Fig. 2d, e).

For cells longer than 1.50 μm with only a single visible copy of the construct, the cellular position was classified as being closer to either the mid-cell (Fig. 2b) or the cell pole (Fig. 2c). In the absence of SbcCD-induced DNA double-strand breaks, the construct was found to be located in the two positions in approximately equal proportions (Fig. 3b). Notably, inducing SbcCD expression for 90 min in a strain containing a DNA palindrome resulted in the construct being located at the cell pole in most cells (89.2%). The slight bias towards being located at the cell pole (63.5%) at T_0 is probably caused by leaky expression of SbcCD from the inducible promoter (P_{BAD})¹² in minimal growth medium. The simplest explanation of this is that SbcCD is specifically cleaving the palindrome located at mid-cell because it possessed a DNA hairpin that formed on the lagging strand of replication. However, other formal possibilities must be considered.

Time-lapse microscopy suggests that the chromosome is cleaved before visible segregation of the sister loci in most cells (data not shown). Therefore, to test the alternative hypothesis that the remaining copy of the palindrome migrated to the cell pole in response to the

induced double-strand break as opposed to the normal process of chromosome segregation, strains were constructed that allowed expression of I-SceI from an inducible promoter with a unique I-SceI cleavage site in place of the palindrome. I-SceI is a site-specific endonuclease and therefore cannot distinguish between the cleavage sites located on the two sister chromosomes. Induction of cleavage by I-SceI for 90 min also resulted in an increase in the number of cells longer than 1.50 μm with a single pair of foci (39.5% after 90 min of induction compared with 8.1% in a control strain lacking a cleavage site), again suggesting degradation of one of the sister chromosomes. The frequency was less than that observed with SbcCD-induced cleavage owing to a more dramatic increase in the cells with no visible foci, presumably because of instances of cleavage and degradation of all chromosomes. Unlike SbcCD-mediated palindrome cleavage, however, the position of the remaining copy of the construct was evenly distributed between mid-cell and cell pole positions (51.6% and 48.4%, respectively; Fig. 3b). This result argues against the possibility of the intact copy of the construct migrating to the cell pole in response to a DNA double-strand break. Furthermore, if cell division was inhibited by treatment with cephalixin to allow multiple rounds of replication and segregation within cells, induction of cleavage by SbcCD did not necessarily result in segregation to a cell pole (Fig. 4a, b).

A final possibility is that SbcCD can only cleave a hairpin when it is located at mid-cell. Consistent with this argument, SbcCD–green-fluorescent protein (GFP) localizes at mid-cell in 54% of the population, with the remainder possessing two foci at the one-quarter and three-quarter positions of the cell¹³. However, its co-localization with the replisomes makes it improbable that SbcCD is able to distinguish the two copies of the palindrome on the basis of their cellular location, considering that segregation of the two replicated sister loci is itself progressive from the replisomes. Also, the frequency of cleavage (more than 36% total DNA as measured on gels⁸ and 52.8% of chromosomes as measured here by the frequency of chromosome degradation in cephalixin-induced filaments; Fig. 4c) is inconsistent with this possibility, which predicts a maximum of 25% cleavage per cell cycle if segregation of the leading and lagging strands is random.

Our data therefore provide strong evidence that segregation of the two sister chromosomes in *E. coli* is not random and results in the placement of the two lagging strands at the cell quarters flanking the mid-cell (Fig. 1Aa). Non-random segregation of sister chromosomes has been observed previously, most notably in mammalian stem cells. This led to the proposal of the immortal strand hypothesis¹⁴, which suggests that stem cells preferentially retain the grandmother DNA template strands to limit the accumulation of replication-induced

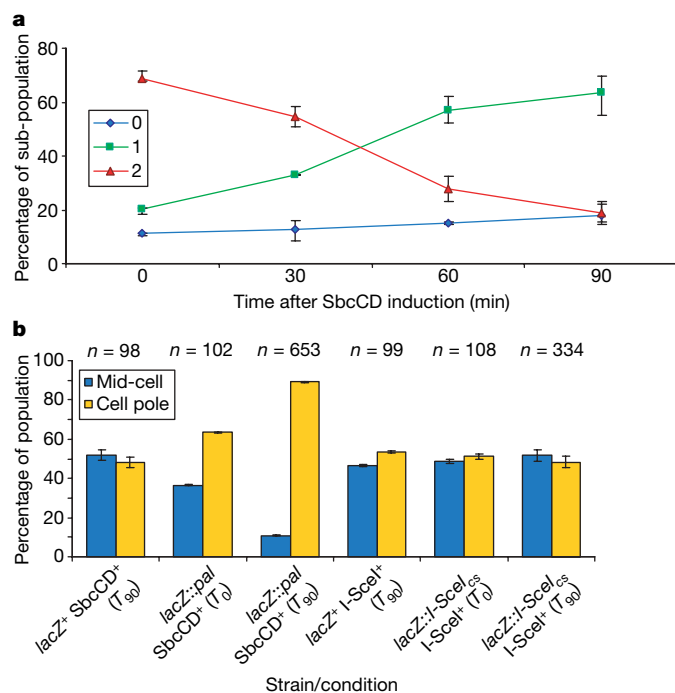


Figure 3 | Construct degradation and localization. **a**, Number of visibly segregated copies of the construct (0, 1, 2) in a *recA*[−]–palindrome-containing strain as a function of time of SbcCD induction (error bars, range; $n \geq 500$ for all time points). **b**, Location of construct in cells longer than 1.50 μm with a single segregated copy of the construct; error bars, range; pal, palindrome.

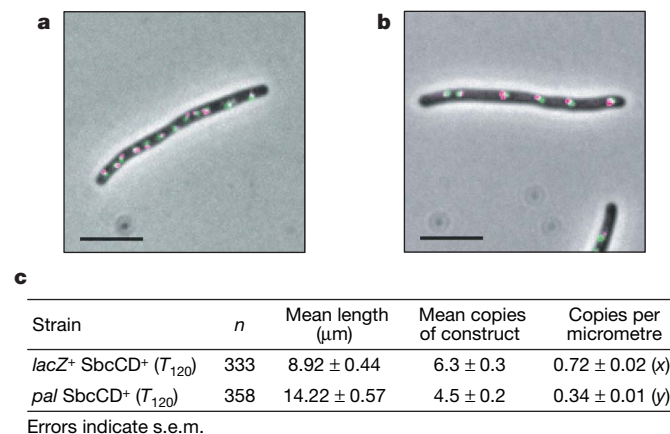


Figure 4 | Induction of SbcCD in cephalixin-induced filaments.

Cephalixin-treated cell without (**a**) and with (**b**) the palindrome after 120 min of SbcCD induction. The CFP signal is pseudocoloured green, YFP is magenta. Overlapping CFP and YFP signals appear white. Scale bars, 5 μm . **c**, Analysis of the consequences of SbcCD expression in cephalixin-treated cells. Frequency of cleavage was calculated as $1 - y/x$.

mutations. The immortal strand hypothesis is controversial^{15–17}, however, and other explanations for the observed asymmetric segregation, such as epigenetic markers, have been proposed¹⁸. These hypotheses require a mechanism of distinguishing sister chromatids. It has recently been postulated that such a discrimination could be made on the basis of whether the centromeric region of the chromosomes were replicated on the leading or lagging strand¹⁹. Here we have shown that a similar distinction is made by *E. coli*.

Our data do not provide a mechanistic understanding of this aspect of chromosome segregation. This is not surprising considering that the mechanism underlying chromosome segregation in *E. coli* is still unknown. The velocity at which the newly replicated origin regions are segregated indicates that chromosome segregation is an active process²⁰, and a centromere-like sequence (*migS*) has been identified²¹. Because later replicated regions of the chromosome do not migrate as rapidly, it has been proposed that segregation is a two-step process instigated by the active positioning of the origins²². Recent evidence suggests a role for the structural maintenance of chromosome protein MukB²³ in this initial step. It has also been suggested that the actin-like cytoskeleton protein MreB has a role in chromosome segregation^{24,25}, although this remains controversial²⁶.

One possibility is that the pattern of segregation that we have observed is a consequence of the structure of the two replisomes at the replication factory²⁷. In *E. coli* the leading and lagging strand polymerases are physically linked in a structure known as the replisome. In the replisome, the lagging strand template is looped to allow the two polymerases to translocate in the same direction¹. Although this loop is small, it may be enough to constrain the lagging strands physically to the replisome while allowing the leading strands to segregate to the cell poles. This would fit with the replication factory model, where the two replisomes are located at mid-cell. It is also plausible that it is advantageous for the cell to keep newly replicated lagging strands at mid-cell for further processing²⁸, and that this local requirement is reflected in global organization. Recent evidence, however, argues against the existence of a replication factory in *E. coli* and implies that the replisomes simply track along the DNA²⁹. If this is indeed the case, then it would suggest that newly replicated DNA follows a nucleoid organization that is established at the initial stages of the segregation process.

METHODS SUMMARY

Strains and plasmids. Strains were BW27784 derivatives allowing homogeneous expression from P_{BAD} ¹². Mutations were introduced by either P1 transduction or plasmid-mediated gene replacement⁸. Insertions and deletions were confirmed by PCR. *recA*[−] mutants were confirmed as ultraviolet light sensitive. Full details of strains and plasmids used can be found in Supplementary Information.

Microscopy. Strains were grown at 37 °C in M9 minimal medium (with 0.2% glycerol) supplemented with 100 ng ml^{−1} anhydrotetracycline to prevent operator-bound TetR–YFP from blocking replication³⁰. Unless otherwise stated, live cells were mounted on a bed of 1% agarose–H₂O for viewing under the microscope. To induce expression from P_{BAD} , cells were grown until the optical density at 600 nm ($OD_{600\text{ nm}}$) = 1.0 in minimal medium supplemented with 0.2% glucose, before diluting to $OD_{600\text{ nm}}$ = 0.2 in minimal medium supplemented with 0.001% arabinose (T_0). To inhibit cell division, cells were grown for 90 min in the presence of 10 $\mu\text{g ml}^{-1}$ cephalixin before inducing SbcCD expression by diluting to $OD_{600\text{ nm}}$ = 0.2 in minimal medium (with 0.001% arabinose) supplemented with cephalixin. SbcCD was induced for 120 min before image acquisition. To prevent cross-signal from the plasmid, nucleoids were visualized in strains DL3339 and DL3340. Nucleoids were visualized either by staining with DAPI (1 $\mu\text{g ml}^{-1}$) or mounting live cells on 27% bovine gelatine–M9 minimal medium.

Images were acquired at a resolution of 0.129 μm per pixel using a Zeiss Axiovert 200 fluorescence microscope equipped with a Photometrics coolSNAP HQ CCD (charge-coupled device) camera and the acquisition software MetaMorph 6.3r2. CFP and YFP images were subject to three-dimensional adaptive point-spread function (blind) deconvolution using Autodeblur and Autovisualize v9.3 before analysis. Cell lengths were measured using the ‘fibre length’ measurement of MetaMorph 6.3r2. Images were pseudocoloured using MetaMorph 6.3r2. The range indicates the result of two independent experiments.

Received 16 May; accepted 23 July 2008.

- McInerney, P., Johnson, A., Katz, F. & O'Donnell, M. Characterization of a triple DNA polymerase replisome. *Mol. Cell* **27**, 527–538 (2007).
- Wang, X., Liu, X., Possoz, C. & Sherratt, D. J. The two *Escherichia coli* chromosome arms locate to separate cell halves. *Genes Dev.* **20**, 1727–1731 (2006).
- Nielsen, H. J. *et al.* The *Escherichia coli* chromosome is organized with the left and right chromosome arms in separate cell halves. *Mol. Microbiol.* **62**, 331–338 (2006).
- Thanbichler, M. & Shapiro, L. Getting organized—how bacterial cells move proteins and DNA. *Nature Rev. Microbiol.* **6**, 28–40 (2008).
- Wang, X., Possoz, C. & Sherratt, D. J. Dancing around the divisome: asymmetric chromosome segregation in *Escherichia coli*. *Genes Dev.* **19**, 2367–2377 (2005).
- Woldringh, C. L. & Nanninga, N. Structural and physical aspects of bacterial chromosome segregation. *J. Struct. Biol.* **156**, 273–283 (2006).
- Rocha, E. P. *et al.* A strand-specific model for chromosome segregation in bacteria. *Mol. Microbiol.* **49**, 895–903 (2003).
- Eykelenboom, J. K., Blackwood, J. K., Okely, E. & Leach, D. R. SbcCD causes a double-strand break at a DNA palindrome in the *Escherichia coli* chromosome. *Mol. Cell* **29**, 644–651 (2008).
- Pinder, D. J., Blake, C. E., Lindsey, J. C. & Leach, D. R. Replication strand preference for deletions associated with DNA palindromes. *Mol. Microbiol.* **28**, 719–727 (1998).
- Skarstad, K. & Boye, E. Degradation of individual chromosomes in *recA* mutants of *Escherichia coli*. *J. Bacteriol.* **175**, 5505–5509 (1993).
- Mason, D. J. & Powelson, D. M. Nuclear division as observed in live bacteria by a new technique. *J. Bacteriol.* **71**, 474–479 (1956).
- Khlebnikov, A. *et al.* Homogeneous expression of the P_{BAD} promoter in *Escherichia coli* by constitutive expression of the low-affinity high-capacity AraE transporter. *Microbiology* **147**, 3241–3247 (2001).
- Darmon, E. *et al.* SbcCD regulation and localization in *Escherichia coli*. *J. Bacteriol.* **189**, 6686–6694 (2007).
- Cairns, J. Mutation selection and the natural history of cancer. *Nature* **255**, 197–200 (1975).
- Rando, T. A. The immortal strand hypothesis: segregation and reconstruction. *Cell* **129**, 1239–1243 (2007).
- Kiel, M. J. *et al.* Haematopoietic stem cells do not asymmetrically segregate chromosomes or retain BrdU. *Nature* **449**, 238–242 (2007).
- Waghmare, S. K. *et al.* Quantitative proliferation dynamics and random chromosome segregation of hair follicle stem cells. *EMBO J.* **27**, 1309–1320 (2008).
- Lansdorp, P. M. Immortal strands? Give me a break. *Cell* **129**, 1244–1247 (2007).
- Lew, D. J., Burke, D. J. & Dutta, A. The immortal strand hypothesis: how could it work? *Cell* **133**, 21–23 (2008).
- Niki, H., Yamaichi, Y. & Hiraga, S. Dynamic organization of chromosomal DNA in *Escherichia coli*. *Genes Dev.* **14**, 212–223 (2000).
- Yamaichi, Y. & Niki, H. *migS*, a cis-acting site that affects bipolar positioning of *oriC* on the *Escherichia coli* chromosome. *EMBO J.* **23**, 221–233 (2004).
- Thanbichler, M. & Shapiro, L. Chromosome organization and segregation in bacteria. *J. Struct. Biol.* **156**, 292–303 (2006).
- Danilova, O. *et al.* MukB colocalizes with the *oriC* region and is required for organization of the two *Escherichia coli* chromosome arms into separate cell halves. *Mol. Microbiol.* **65**, 1485–1492 (2007).
- Gitai, Z. *et al.* MreB actin-mediated segregation of a specific region of a bacterial chromosome. *Cell* **120**, 329–341 (2005).
- Kruse, T. *et al.* Actin homolog MreB and RNA polymerase interact and are both required for chromosome segregation in *Escherichia coli*. *Genes Dev.* **20**, 113–124 (2006).
- Karczmarek, A. *et al.* DNA and origin region segregation are not affected by the transition from rod to sphere after inhibition of *Escherichia coli* MreB by A22. *Mol. Microbiol.* **65**, 51–63 (2007).
- Lemon, K. P. & Grossman, A. D. Localization of bacterial DNA polymerase: evidence for a factory model of replication. *Science* **282**, 1516–1519 (1998).
- Rossi, M. L., Purohit, V., Brandt, P. D. & Bambara, R. A. Lagging strand replication proteins in genome stability and DNA repair. *Chem. Rev.* **106**, 453–473 (2006).
- Reyes-Lamothe, R., Possoz, C., Danilova, O. & Sherratt, D. J. Independent positioning and action of *Escherichia coli* replisomes in live cells. *Cell* **133**, 90–102 (2008).
- Possoz, C., Filipe, S. R., Grainge, I. & Sherratt, D. J. Tracking of controlled *Escherichia coli* replication fork stalling and restart at repressor-bound DNA *in vivo*. *EMBO J.* **25**, 2596–2604 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank D. Sherratt for the gifts of plasmids pWX6, pLau43 and pLau44. We also thank E. Darmon and J. Blackwood for reading the manuscript. This work was supported by the Medical Research Council.

Author Contributions M.A.W. and D.R.F.L. conceived and designed the experiments; M.A.W. constructed all strains and plasmids apart from pDL1625, pDL1709 and pDL2542, which were constructed by J.K.E., E.W. and M.A.L.-V., respectively; M.A.W. performed the experiments; M.A.W. and D.R.F.L. analysed the data and wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to D.R.F.L. (D.Leach@ed.ac.uk).

Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast

Lyris M. F. de Godoy^{1*}, Jesper V. Olsen^{1*}, Jürgen Cox^{1*}, Michael L. Nielsen^{1*}, Nina C. Hubner¹, Florian Fröhlich², Tobias C. Walther² & Matthias Mann¹

Mass spectrometry is a powerful technology for the analysis of large numbers of endogenous proteins^{1,2}. However, the analytical challenges associated with comprehensive identification and relative quantification of cellular proteomes have so far appeared to be insurmountable³. Here, using advances in computational proteomics, instrument performance and sample preparation strategies, we compare protein levels of essentially all endogenous proteins in haploid yeast cells to their diploid counterparts. Our analysis spans more than four orders of magnitude in protein abundance with no discrimination against membrane or low level regulatory proteins. Stable-isotope labelling by amino acids in cell culture (SILAC) quantification^{4,5} was very accurate across the proteome, as demonstrated by one-to-one ratios of most yeast proteins. Key members of the pheromone pathway were specific to haploid yeast but others were unaltered, suggesting an efficient control mechanism of the mating response. Several retrotransposon-associated proteins were specific to haploid yeast. Gene ontology analysis pinpointed a significant change for cell wall components in agreement with geometrical considerations: diploid cells have twice the volume but not twice the surface area of haploid cells. Transcriptome levels agreed poorly with proteome changes overall. However, after filtering out low confidence microarray measurements, messenger RNA changes and SILAC ratios correlated very well for pheromone pathway components. Systems-wide, precise quantification directly at the protein level opens up new perspectives in post-genomics and systems biology.

Yeast launched the genome era⁶ and continues to be an informative model system for genomic and post-genomics technologies. It has also been a fruitful testing ground for mass spectrometry (MS)-based proteomics^{7–10}. Repositories of yeast proteomics experiments contain about 4,000 proteins, albeit with varying confidence of identification¹¹. Previously, we established that half of the yeast proteome could be detected with very high stringency by MS in a single experiment¹². The phosphoproteome of pheromone signalling has already been investigated by a SILAC experiment¹³. Until now, no strategies have been described to comprehensively identify, much less to comprehensively quantify, two states of the yeast proteome against each other in a single experiment.

To develop methods for proteome-wide quantification, we metabolically labelled haploid and diploid yeast with arginine and lysine SILAC. We investigated three strategies to achieve deep coverage of the yeast proteome: extensive fractionation of proteins; fractionation of digested peptides; and accumulating and sequencing distinct mass ranges of peptides (Fig. 1, Methods). The second strategy, combining in-solution digest with peptide separation by isoelectric focusing, yielded the most proteins (3,987) and is by far the simplest.

Together, we identified 4,399 proteins with 99% certainty (Supplementary Table 4). Unambiguous identification only requires

a few peptides per protein; however, on average we covered 32% of each protein sequence.

Previously, expressed yeast genes were detected by a fused tandem affinity tag (TAP)¹⁴ or green fluorescent protein (GFP) tag¹⁵ in genome-wide experiments (Fig. 2a and Table 1) and our data overlaps 89% with each of these tagging approaches. In addition, MS identified 510 proteins exclusively, including proteins in which the tag interferes with function, such as tail-anchored membrane proteins and proteins requiring carboxy-terminal modifications. As judged by MS, several hundred proteins previously reported at less than 50 copies per cell were part of different abundance classes over the whole dynamic range (Supplementary Fig. 5). Our data set is not biased against low-abundance proteins (Fig. 2b) or membrane proteins (30.9% of all proteins detected and 29.4% of the genome). Only 6% of yeast open reading frames (ORFs) were detected by both tagging methods but not by MS (Fig. 2a). This is less than the discrepancy between the tagging methods and includes 12 proteins that are inaccessible to MS due to a lack of appropriate tryptic or LysC cleavage sites, 33 proteins with overlapping genes (which we only counted as single identifications), 11 that have been removed from the database during the last three years, 8 dubious genes and 78 proteins for which no western blot quantification had been possible. Thus, of the accessible proteome, at most a few per cent of proteins are not detected. High-resolution data from the orbitrap instrument combined with efficient computational strategies led to very high peptide mass accuracy (average absolute mass deviation of 590 p.p.b.) and to very high identification rates for mass spectrometric peptide fragmentation (>53% on SILAC peptide pairs, Fig. 1d and Methods), contributing to the identification of essentially the entire yeast proteome expressed in log-phase cells.

Next, we determined the fold change of SILAC peptide pairs for relative proteome quantification between haploid and diploid yeast cells. In arginine and lysine double-labelled populations, we noticed that the proteomes were substantially different due to the presence of different sets of auxotrophic markers in the haploid and diploid strains (Supplementary Fig. 6). We therefore based our quantitative analysis on the lysine-labelled haploid S288C yeast strain and compared it to an isogenic diploid strain (Fig. 1b, c and Methods). A total of 1,788,451 SILAC peptide pairs were identified and quantified (median of 32 pairs per protein). Figure 3a and Supplementary Tables 6 and 7 show the ratios of all 4,033 quantified proteins and peptides from the lysine-labelling experiments. We achieved very high quantification accuracy, with 97.3% of the proteome changing less than 50% in abundance between haploid and diploid cells. Quantification after fractionation of digested peptides (Fig. 1b) showed excellent reproducibility ($R = 0.84$ on average; Supplementary Fig. 7). One-hundred-and-ninety-six proteins changed significantly ($P < 0.001$), and we confirmed the regulation of 29 of

¹Proteomics and Signal Transduction, and ²Organelle Architecture and Dynamics, Max-Planck-Institute for Biochemistry, Am Klopferspitz 18, D-82152 Martinsried, Germany.

*These authors contributed equally to this work.

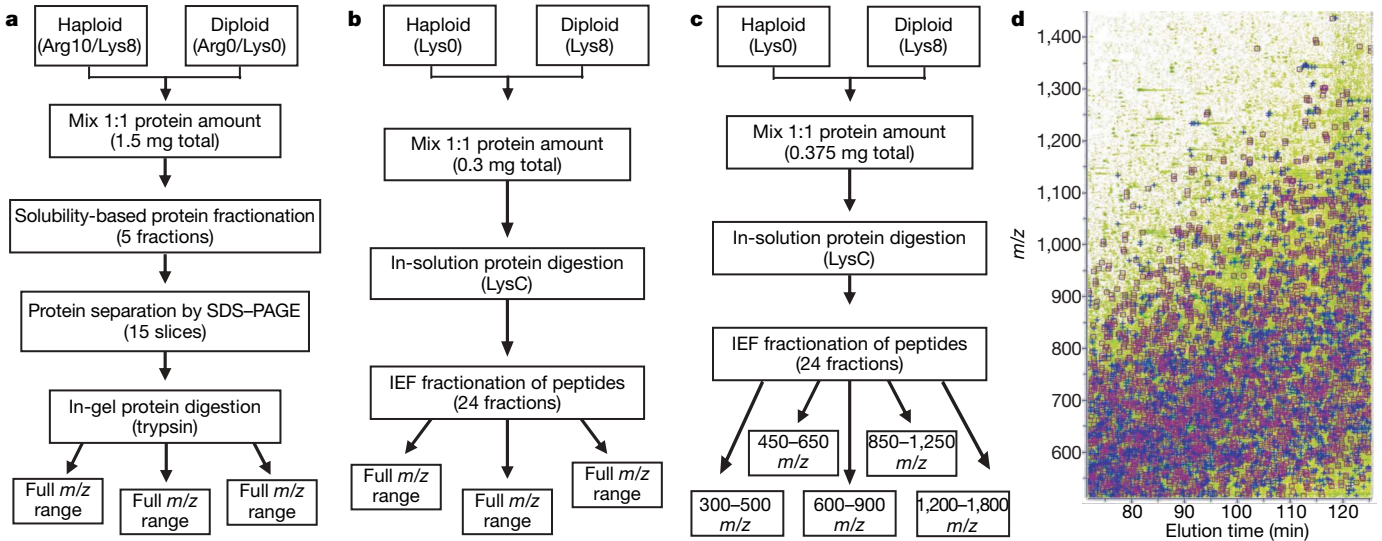


Figure 1 | Three strategies for in-depth quantification of the yeast proteome by SILAC labelling and high-resolution mass spectrometry. **a**, Arginine and lysine SILAC labelling of haploid and diploid yeast. Arg10 is [¹³C₆,¹⁵N₄]L-arginine, Lys8 is [¹³C₆,¹⁵N₂]L-lysine, and Arg0 and Lys0 are the normal, non-substituted amino acids. Extensive fractionation followed by tryptic digestion and one-dimensional gel electrophoresis as well as online LC-MS/MS on a hybrid linear ion trap-orbitrap instrument yielded, through triplicate measurements, 3,639 identified proteins at high stringency using the MaxQuant algorithms (J.C. and M.M., submitted; Supplementary Table 1). **b**, Lysine SILAC labelling of haploid and diploid yeast. Triplicate measurements of in-solution digestion with endoprotease

LysC followed by isoelectric focusing into 24 fractions and online LC-MS/MS resulted in a proteome of 3,987 proteins (Supplementary Table 2). **c**, Same as **b** except that each isoelectric fraction is analysed five times with ion accumulation of a narrow *m/z* range for higher dynamic range. The signal-to-noise ratio and dynamic range improved by about a factor of five (Supplementary Fig. 1) and 3,779 proteins were identified (Supplementary Table 3). **d**, Typical contour plot of a single LC-MS/MS run. Peptide pairs eluting from the column (green) were automatically fragmented (blue crosses) and more than 60% of sequencing events on SILAC pairs resulted in successful identification (purple boxes).

the top-regulated ones by western blot against either the fused TAP or GFP tag from the systematic collection¹⁴ (Supplementary Fig. 8). All ratios were in the same direction as that observed by MS-based proteomics. Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway and Gene Ontology analysis (Supplementary Table 8) highlighted lysine biosynthesis as being upregulated in diploid cells ($P = 5 \times 10^{-6}$). This is due to heterozygosity for *LYS2/lys2* and illustrates the ability of proteome-wide quantification to pinpoint altered metabolic pathways (Supplementary Fig. 9a, c).

Pheromone signalling is required for mating of haploid cells and is absent from diploid cells¹⁶. The top ten haploid-specific proteins as determined by SILAC are components or transcriptional targets of pheromone signalling (Supplementary Table 9). Surprisingly, not all of its members are regulated equally (Fig. 3b). Key components of the signal transduction pathway and output factors were absent from diploid cells: the pheromone receptor (Ste2), the signal transducing G protein (consisting of Ste4, Ste18 and Gpa1), the mitogen-activated protein kinase (MAPK) scaffold protein Ste5, the MAPK Fus3 and the output transcription factor Ste12. In contrast, the

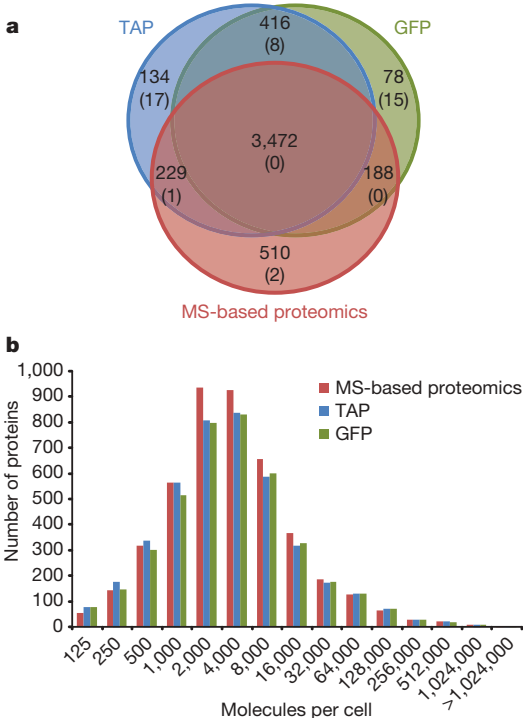


Figure 2 | Proteome coverage. **a**, Comparison of coverage of MS-based proteomics with GFP- and TAP-tagging methods^{14,15}. Numbers are the identified proteins by each method and, in parentheses, the number of dubious open reading frames (ORFs). **b**, Identified proteins per copy number bin for MS-based proteomics and the two tagging approaches. Copy numbers were estimated by correlation between summed peptide intensity per protein and the quantitative western blotting data¹⁴ (Methods).

Table 1 Yeast ORFs identified by SILAC-based quantitative proteomics				
	Number of ORFs	TAP	GFP	nanoLC-MS
Total yeast ORFs	6,608	4,251	4,154	4,399
Characterized yeast ORFs	4,666	3,629	3,581	3,824
Uncharacterized yeast ORFs	1,128	581	539	572
Dubious yeast ORFs	814	26 (3%)	23 (3%)	3 (<1%)
Not present in ORF database		15	11	0

Comparative sequencing shows that 814 of the 6,608 yeast ORFs are never expressed (dubious ORFs, <http://www.yeastgenome.org>). Of these only six were identified in this experiment and three were validated by SILAC-assisted *de novo* sequencing of several peptides (Supplementary Table 5 and Supplementary Figs 2–4). Two of the three validated ones were reclassified as genuine yeast genes during writing of this manuscript (YGL041W-A and YPR170W-B). This leaves three potential false-positives (0.37% of 815) and suggests that our estimate of a false-positive identification rate of maximally 1% is conservative.

MAPKKKK Ste20, the MAPKKK Ste11 and the MAPKK Ste7 remained unchanged. For some of these kinases, such as Ste7 or Ste11, this is readily explained because they fulfil another function in the osmolarity-sensing and filamentous growth pathway¹⁷. For other proteins, such as the Far3/7/8/11 protein complex that mediates one pathway of cell cycle arrest during the pheromone response, this is unexpected and might indicate that they have another function in haploid cells. This suggests another repressive function of Far3 during the cell cycle. Consistently, its inactivation results in faster growth of haploid cells¹⁸.

The proteins encoded by retrotransposons Ty1 and Ty2 are about ten times more abundant in haploid cells, consistent with regulation of specific Ty mRNAs by pheromone signalling in haploid cells and repression in diploid cells by the MAT α / α transcription factor^{19,20}. We also found the Ty1 transcription activator Tec1 to be eight times more expressed in haploid cells. Little is known about the evolutionary advantage of restricting retrotransposition to haploid cells, but because most wild-type cells are diploid, the repression of transposition in these cells might be used to minimize the spread of detrimental effects through the population.

Cell wall components were statistically significantly reduced in diploid cells ($P = 2.7 \times 10^{-9}$; Supplementary Table 8). At first glance, this is surprising because diploid cells are on average twice as large as haploid cells and also have more cell wall. However, larger cells need less surface components in relation to 'bulk' proteins, and the observed downregulation (0.77) is very close to what would be expected from geometrical considerations: a sphere of double volume has $2^{2/3}$ the surface and thus should have $2^{2/3}/2 = 0.79$ the amount of surface proteins after normalization for the doubled volume. The list of differentially expressed factors also contains a number of uncharacterized genes, which can be mined for haploid-specific functions.

A longstanding question in functional genomics is to what extent changes in mRNA levels lead to changes of the active agents in the cell, the proteins²¹. Overall correlation of mRNA²² and protein changes was poor ($R = 0.24$) and there were large populations of genes with mRNA but no protein change (Fig. 4a). However, after we filtered out low-level microarray signals (Supplementary Fig. 10), the correlation improved to 0.46 (Fig. 4b). Several of the remaining, discordant mRNA changes seem to be technical artefacts. For example, *INO1*, the protein level of which did not change, is the only representative of several co-regulated genes (for example, *CHO1* and

CHO2) that was found upregulated by microarray analysis. *CTS1*, which was downregulated according to microarray analysis, was upregulated when measured by SILAC and western blot. Several lysine biosynthesis pathway genes seem to be regulated at the protein but not the mRNA level (magenta in Fig. 4b). However, this is due to use of lysine auxotrophs in the MS but not the microarray experiments. Among genes only found upregulated by proteomics (blue in Fig. 4b), cell wall proteins were highly overrepresented ($P = 7.7 \times 10^{-8}$, see Methods). This could be due to the microarray experiment not detecting slight expression changes for this class of proteins. Strongly regulated genes in both data sets were mainly components of the pheromone response. Here, correlation between mRNA and protein changes was high ($R = 0.68$; Fig. 4c). However, actual fold changes determined by microarrays deviated considerably from the values provided by the SILAC quantification (Supplementary Table 10). This is probably due to technical differences (that is, microarray measurements are not strictly quantitative) combined with the fact that the level of mRNA change may not directly be translated into a change of protein level.

In summary, a combination of SILAC labelling, high-resolution MS and sophisticated computational proteomics allows accurate quantitative analysis of an entire proteome. Among several tested strategies, in-solution digest of unfractionated cell lysate followed by simple isoelectric focusing of the peptides proved most powerful.

Key advantages of MS-based proteomics are the ability to measure endogenous rather than tagged versions of proteins, which may have altered expression levels, and to quantify the entire proteome from one sample. Our comparison of the proteome with the transcriptome highlights several crucial points for systems-wide analysis. First, proteomics can directly measure small changes in the amounts of proteins, which might have important effects in the cell. Second, it shows that the relationship between mRNA and protein levels depends on the proteins investigated. This effect is likely to be even more notable in mammalian proteomes, which compared to yeast are more complex and subject to more post-transcriptional control. A mammalian cell is commonly thought to express 10,000 gene products, which would only be two to three times the number of genes expressed in yeast. Thus, we predict that essentially complete mammalian proteomes—with at least one representative protein per expressed gene—will be feasible with refined versions of our strategy²³. The next challenge will then be proteome-wide identification of functionally important isoforms and modifications.

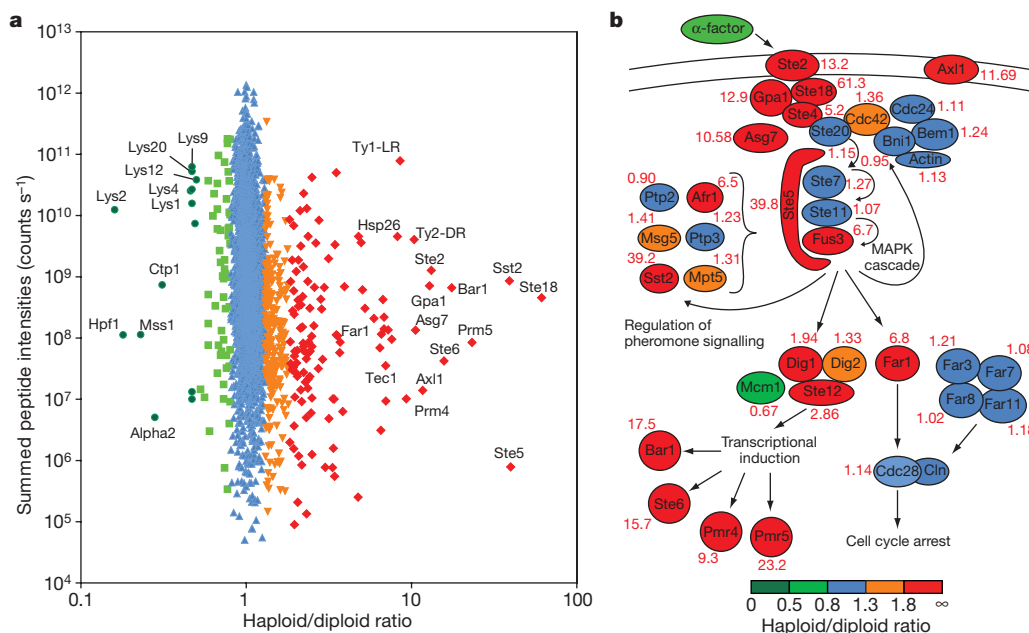


Figure 3 | Quantitative differences between the haploid and diploid yeast proteome. **a**, Overall fold change for the yeast proteome. **b**, Members of the yeast pheromone response are colour-coded according to fold change. The diploid to haploid ratio as determined by SILAC is indicated for each protein. Figure is adapted from ref. 13.

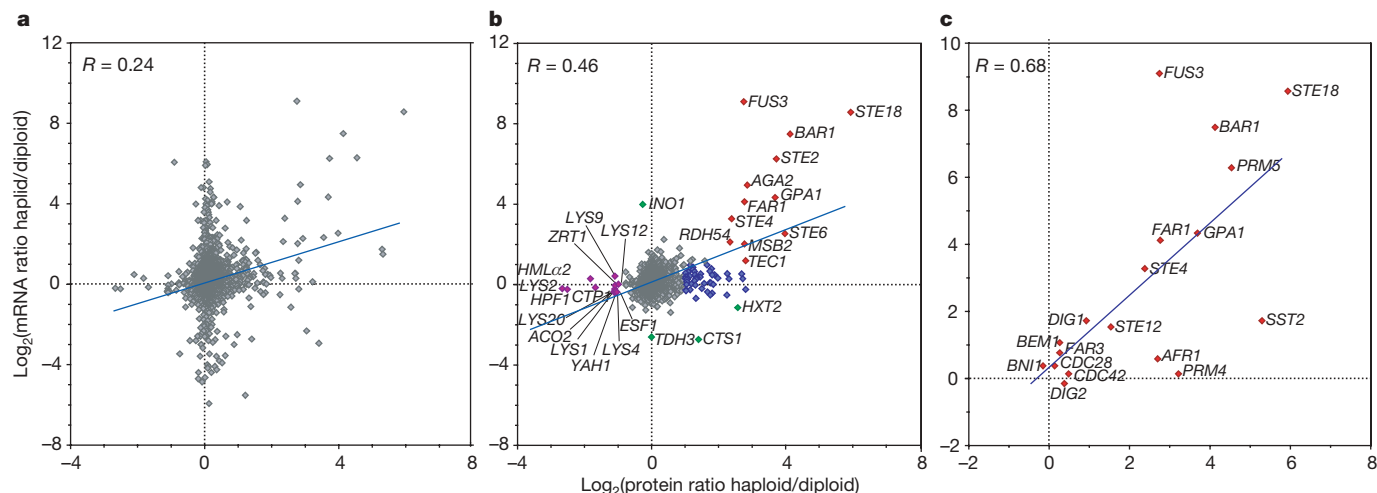


Figure 4 | Proteome and transcriptome changes of haploid versus diploid yeast. **a**, The overall correlation between protein and mRNA changes is poor ($R = 0.24$). **b**, After filtering out low mRNA signals, the data from **a** correlates better (Supplementary Fig. 10). Red, significantly upregulated as mRNA and protein; blue, significantly upregulated as protein; green, significantly

changed as mRNA; and magenta, significantly downregulated as protein. **c**, Proteins involved in pheromone response (Fig. 3b) are co-regulated at mRNA and protein levels but actual protein ratios cannot be accurately predicted from changes of mRNA levels.

METHODS SUMMARY

Yeast diploid and haploid strains were SILAC-labelled as described¹³ with [¹³C₆/¹⁵N₂]-L-lysine-and/or [¹³C₆/¹⁵N₄]-L-arginine. The diploid yeast strain TWY 809 was generated by crossing the wild-type BY4741 and BY4742. The haploid strain for lysine labelling was generated by sporulation of BY4743 and selection for the lysine auxotroph, MATa cells. Yeast cells were lysed, mixed 1:1, fractionated by SDS-PAGE and in-gel digested with trypsin as described previously¹². Alternatively, after mixing, proteins were digested in-solution by the endoproteinase LysC and the resulting peptide mixtures were fractionated by peptide isoelectric focusing. Each fraction was subsequently analysed by online liquid chromatography–tandem mass spectrometry (LC–MS/MS). All LC–MS/MS experiments were performed on an LTQ-Orbitrap (Thermo Fisher Scientific) mass spectrometer connected to an Agilent 1200 nanoHPLC system by means of a nanoelectrospray source (Proxeon Biosystems). MS full scans were acquired in the Orbitrap analyser using internal lock mass recalibration in real-time²⁴ whereas tandem mass spectra were simultaneously recorded in the linear ion trap. Peptides were identified from MS/MS spectra by searching them against the yeast ORF database (Stanford University) using the Mascot search algorithm²⁵ (<http://www.matrixscience.com>), and all SILAC pairs were quantified by MaxQuant (J.C. and M.M., submitted). For several of the top-regulated proteins, GFP- or TAP-tagged haploid and diploid strains were generated and the regulation was confirmed by western blot.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 30 May; accepted 12 August 2008.

Published online 28 September 2008.

1. Aebersold, R. & Mann, M. Mass spectrometry-based proteomics. *Nature* **422**, 198–207 (2003).
2. Cravatt, B. F., Simon, G. M. & Yates, J. R. III. The biological impact of mass-spectrometry-based proteomics. *Nature* **450**, 991–1000 (2007).
3. Malmstrom, J., Lee, H. & Aebersold, R. Advances in proteomic workflows for systems biology. *Curr. Opin. Biotechnol.* **18**, 378–384 (2007).
4. Ong, S. E. *et al.* Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell. Proteomics* **1**, 376–386 (2002).
5. Mann, M. Functional and quantitative proteomics using SILAC. *Nature Rev. Mol. Cell Biol.* **7**, 952–958 (2006).
6. Goffeau, A. *et al.* Life with 6000 genes. *Science* **274**, 563–567 (1996).
7. Shevchenko, A. *et al.* Linking genome and proteome by mass spectrometry: large-scale identification of yeast proteins from two dimensional gels. *Proc. Natl Acad. Sci. USA* **93**, 14440–14445 (1996).
8. Figeys, D. *et al.* Protein identification by solid phase microextraction-capillary zone electrophoresis-microelectrospray-tandem mass spectrometry. *Nature Biotechnol.* **14**, 1579–1583 (1996).
9. Washburn, M. P., Wolters, D. & Yates, J. R. III. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nature Biotechnol.* **19**, 242–247 (2001).

10. Peng, J. *et al.* Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC–MS/MS) for large-scale protein analysis: the yeast proteome. *J. Proteome Res.* **2**, 43–50 (2003).
11. King, N. L. *et al.* Analysis of the *Saccharomyces cerevisiae* proteome with PeptideAtlas. *Genome Biol.* **7**, R106 (2006).
12. de Godoy, L. M. *et al.* Status of complete proteome analysis by mass spectrometry: SILAC labeled yeast as a model system. *Genome Biol.* **7**, R50 (2006).
13. Gruhler, A. *et al.* Quantitative phosphoproteomics applied to the yeast pheromone signaling pathway. *Mol. Cell. Proteomics* **4**, 310–327 (2005).
14. Ghaemmaghami, S. *et al.* Global analysis of protein expression in yeast. *Nature* **425**, 737–741 (2003).
15. Huh, W. K. *et al.* Global analysis of protein localization in budding yeast. *Nature* **425**, 686–691 (2003).
16. Dohlman, H. G. & Slessareva, J. E. Pheromone signaling pathways in yeast. *Sci. STKE* **2006**, cm6 (2006).
17. Schwartz, M. A. & Madhani, H. D. Principles of MAP kinase signaling specificity in *Saccharomyces cerevisiae*. *Annu. Rev. Genet.* **38**, 725–748 (2004).
18. Blanc, V. M. & Adams, J. Evolution in *Saccharomyces cerevisiae*: identification of mutations increasing fitness in laboratory populations. *Genetics* **165**, 975–983 (2003).
19. Company, M., Errede, B. & Ty, A. A cell-type-specific regulatory sequence is a recognition element for a constitutive binding factor. *Mol. Cell. Biol.* **8**, 5299–5309 (1988).
20. Ke, N., Irwin, P. A. & Voytas, D. F. The pheromone response pathway activates transcription of Ty5 retrotransposons located within silent chromatin of *Saccharomyces cerevisiae*. *EMBO J.* **16**, 6272–6280 (1997).
21. Tyers, M. & Mann, M. From genomics to proteomics. *Nature* **422**, 193–197 (2003).
22. Galitski, T. *et al.* Ploidy regulation of gene expression. *Science* **285**, 251–254 (1999).
23. Cox, J. & Mann, M. Is proteomics the new genomics? *Cell* **130**, 395–398 (2007).
24. Olsen, J. V. *et al.* Parts per million mass accuracy on an orbitrap mass spectrometer via lock mass injection into a C-trap. *Mol. Cell. Proteomics* **4**, 2010–2021 (2005).
25. Perkins, D. N. *et al.* Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20**, 3551–3567 (1999).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements G. de Souza measured part of the yeast proteome; C. Kumar contributed to bioinformatic analysis, G. Stoehr to proteome analysis. Z. Storchova provided the pGAL-HO plasmid. L.M.F.d.G. thanks D. Bertozzi for support and discussions. The Max-Planck Society and the DC-Thera and Interaction Proteome 6th framework projects of the European Union provided funding; T.C.W. is supported by the Human Frontier Science Program and M.L.N. by the European Molecular Biology Organization (EMBO).

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to T.C.W. (twalther@biochem.mpg.de) or M.M. (mmann@biochem.mpg.de).

METHODS

Generation and SILAC-labelling of haploid and diploid yeast strains. The *Saccharomyces cerevisiae* diploid strain YLG1 was generated by crossing the haploid YAL6B MATa strain¹³ with one of its parental strains, Y15969 MATα (Euroscarf). The diploid yeast strain TWY 809 was generated by crossing the wild-type BY4741 and BY4742. The haploid strain for lysine labelling was generated by sporulation of BY4743 and selection for the lysine auxotroph, MATa cells. The arginine and lysine double SILAC labelling was performed as described¹³, with small modifications. In brief, cells from the haploid YAL6B strain, which has a *LYS1* and *ARG4* gene deletions and is therefore a double auxotroph for lysine and arginine, and the diploid YLG1 strain were grown in YNB liquid medium containing either 20 mg l⁻¹ [¹³C₆/¹⁵N₂]L-lysine (Lys8) and 5 mg l⁻¹ [¹³C₆/¹⁵N₄]L-arginine (Arg10; Isotec-Sigma) or 20 mg l⁻¹ L-lysine and 5 mg l⁻¹ L-arginine for ten generations, until they reached log-phase (*D*₆₀₀ 0.7).

Lysis and protein fractionation strategy. Normal and heavy SILAC-labelled yeast cells were collected by centrifugation, resuspended in lysis buffer (150 mM potassium acetate, 2 mM magnesium acetate, 1× protease inhibitor cocktail (Roche), and 20 mM HEPES, pH 7.4) and frozen in liquid N₂. Haploid and diploid frozen cells were mixed 1:1 on the basis of protein amount (as determined by Bradford assay) and mechanically disrupted in a milling device (MM301 Ball Mill, Retsch), with 3 cycles of 3 min at 10 Hz, intercalated by immersion in liquid N₂. All further steps were performed at 4 °C. The extract was allowed to thaw and centrifuged for 4 min at 1,000g. The pellet was collected, washed twice with lysis buffer, resuspended in PBS containing 2% SDS, incubated for 5 min at 65 °C and spun down to remove debris (fraction 1). The sample was centrifuged for 10 min at 20,000g and the resultant pellet washed twice with lysis buffer and resuspended in PBS containing 2% SDS (fraction 2). The supernatant was brought to 60% (NH₄)₂SO₄, incubated for 10 min under rotation to allow protein precipitation, centrifuged for 10 min at 20,000g and the precipitated proteins resuspended in PBS containing 2% SDS (fraction 3). The concentration of (NH₄)₂SO₄ was raised to 80%, the sample processed as before, the precipitated proteins resuspended in PBS containing 2% SDS (fraction 4) and the remaining soluble proteins dialysed against PBS containing 2% SDS (fraction 5).

In-solution digestion. Proteins extracted from lysine-labelled haploid and diploid yeast were reduced for 20 min at room temperature (24 °C) in 1 mM dithiothreitol and then alkylated for 15 min by 5.5 mM iodoacetamide (IAA) at room temperature in the dark. Endoproteinase LysC (Wako) was added 1:50 (w/w) and the lysates were digested overnight at room temperature (12 h). Arginine- and lysine-labelled yeast proteins were digested with LysC in a similar manner, and the resulting peptide mixtures were diluted with Millipore water to achieve a final urea concentration below 2 M. Trypsin (modified sequencing grade, Promega) was added 1:50 (w/w) and digested overnight. Trypsin and LysC activity were quenched by acidification of the reaction mixtures with TFA to ~pH 2.

Peptide isoelectric focusing. In-solution digested peptides (75 µg) were separated according to their isoelectric point using the Agilent 3100 OFFGEL fractionator (Agilent, G3100AA). The system was set up according to the manual of the High Res Kit, pH 3–10 (Agilent, 5188-6424), but strips were exchanged by 24 cm Immobiline DryStrip, pH 3–10 (GE Healthcare, 17-6002-44), and ampholytes were substituted by IPG buffer, pH 3–10 (GE Healthcare, 17-6000-87), used 1:50. Peptides were focused for 50 kilovolt hours (kVh) at a maximum current of 50 µA, maximum voltage of 8,000 V and maximum power of 200 mW into 24 fractions. Each peptide fraction was acidified by adding 3% acetonitrile, 1% trifluoroacetic acid and 0.5% acetic acid, then desalted and concentrated on a reversed-phase C18 StageTip²⁶.

Gel electrophoresis and in-gel digestion. Each lysine- and arginine-labelled yeast protein fraction was boiled in 2× LDS buffer, separated by one-dimensional SDS-PAGE (4–12% Novex mini-gel, Invitrogen) and visualized by colloidal Coomassie staining. The entire protein gel lanes were excised and cut into 20 slices each. Every gel slice was subjected to in-gel digestion with trypsin²⁷. The resulting tryptic peptides were extracted by 30% acetonitrile in 3% TFA, reduced in a Speed Vac, and desalted and concentrated on a reversed-phase C18 StageTip²⁶.

Mass spectrometric analysis. All MS experiments were performed on a nano-flow HPLC system (Agilent Technologies 1200) connected to a hybrid LTQ-orbitrap classic or XL (Thermo Fisher Scientific) equipped with a nanoelectrospray ion source (Proxeon Biosystems) as described²⁴ with a few modifications. In brief, the peptide mixtures were separated in a 15 cm analytical column (75 µm inner diameter) in-house packed with 3-µm C18 beads (Reprosil-AQ Pur, Dr. Maisch) with a 2 h gradient from 5% to 40% acetonitrile in 0.5% acetic acid. The effluent from the HPLC was directly electrosprayed into the mass spectrometer.

The MS instrument was operated in data-dependent mode to automatically switch between full-scan MS and MS/MS acquisition. Survey full-scan MS spectra (from *m/z* 300–2,000) were acquired in the orbitrap with resolution *R* = 60,000 at *m/z* 400 (after accumulation to a 'target value' of 1,000,000 in the linear ion trap).

The ten most intense peptide ions with charge states ≥2 were sequentially isolated to a target value of 5,000 and fragmented in the linear ion trap by collisionally induced dissociation. Fragment ion spectra were recorded with the LTQ detectors 'in parallel' with the orbitrap full-scan detection. For all measurements with the orbitrap detector, a lock-mass ion from ambient air (*m/z* 391.284286, 429.08875 or 445.120025) was used for internal calibration as described²⁴.

For mass range experiments (similar to 'gas-phase fractionation') all samples were analysed using survey scan MS spectra in one of the following mass regions: *m/z* 300–500, *m/z* 450–650, *m/z* 600–900, *m/z* 850–1,250 and *m/z* 1,200–1,800. Resolution, lock mass option, 'target value' and number of intense peptide peaks selected for isolation were identical to full-scan analysis (see below), except for the mass range analysis *m/z* 1,200–1,800 where charge states ≥1 were allowed for isolation. All survey scans were acquired using injection waveforms, which applies a filter on the injection ions and thereby ejects all ions outside of the selected mass range. This ensures optimal dynamic range because the ion trap will only be filled with a population of ions belonging to the mass range of interest.

Identification and quantification of peptides and proteins. The data analysis was performed with the MaxQuant software as described¹³ supported by Mascot as the database search engine for peptide identifications. Peaks in MS scans were determined as three-dimensional hills in the mass-retention time plane. They were then assembled to isotope patterns and SILAC pairs by graph-theoretical methods. MS/MS peak lists were filtered to contain at most six peaks per 100 Da interval and searched by Mascot (Matrix Science) against a concatenated forward and reversed version of the yeast ORF database (*Saccharomyces* Genome Database SGDTM at Stanford University; <http://www.yeastgenome.org>). Protein sequences of common contaminants, for example, human keratins and proteases used, were added to the database. The initial mass tolerance in MS mode was set to 7 p.p.m. and MS/MS mass tolerance was 0.5 Da. Cysteine carbamidomethylation was searched as a fixed modification, whereas N-acetyl protein, N-pyrogutamine and oxidized methionine were searched as variable modifications. Labelled arginine and lysine were specified as fixed or variable modifications, depending on the previous knowledge about the parent ion. The resulting Mascot .dat files were loaded into the MaxQuant software¹³ together with the raw data for further analysis. SILAC peptide and protein quantification was performed automatically with MaxQuant using default settings for parameters. Here, for each SILAC pair the ratio is determined by a robust regression model fitted to all isotopic peaks and all scans that the pair elutes in. SILAC protein ratios are determined as the median of all peptide ratios assigned to the protein. Absolute protein quantification was based on extracted ion chromatograms of contained peptides. To minimize false identifications, all top-scoring peptide assignments made by Mascot were filtered based on previous knowledge of individual peptide mass error, SILAC state and the correct number of lysine and arginine residues specified by the mass difference observed in the full scan between the SILAC partners. Furthermore, peptide assignments were statistically evaluated in a Bayesian model on the basis of sequence length and Mascot score. We accepted peptides and proteins with a false discovery rate of less than 1%, estimated on the basis of the number of accepted reverse hits.

Gene ontology and Pfam domain overrepresentation analysis. *P* values for the overrepresentation of gene ontology categories and protein domain content were based on a Wilcoxon–Mann–Whitney test for the presence–absence pattern of each category and the ratio significance as a continuous value. All *P* values below 0.01 are reported. To determine classes of proteins that show a high protein ratio but only low response on the transcript level, we defined a protein population with a protein ratio above two and a transcript ratio between one-half and two. We looked for enrichment of Gene Ontology terms in this class of proteins compared to the rest by calculating the *P* value according to the Fisher exact test.

SILAC-assisted peptide-sequence-tag searching for ambiguous ORFs. Fragment ion intensities in spectra from 'light' and 'heavy' forms of a SILAC peptide pair are highly correlated. The only difference between their spectra is that C-terminal fragment ions (*y*-ions) are offset by 8.014 Da or other multiples of the difference between normal and heavy labelled amino acids. Extraction of *y*-ions is therefore straightforward and examples are shown in Supplementary Figs 2–4 for each of the three ORFs initially assumed not to be expressed. Searching these SILAC confirmed fragment ions (*y*-ions) in the yeast database as peptide-sequence tags²⁸ unambiguously verified identification of the ORFs.

26. Rappsilber, J., Ishihama, Y. & Mann, M. Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal. Chem.* **75**, 663–670 (2003).

27. Shevchenko, A. et al. Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Anal. Chem.* **68**, 850–858 (1996).

28. Mann, M. & Wilm, M. Error-tolerant identification of peptides in sequence databases by peptide sequence tags. *Anal. Chem.* **66**, 4390–4399 (1994).

Crystal structure of a stable dimer reveals the molecular basis of serpin polymerization

Masayuki Yamasaki¹, Wei Li¹, Daniel J. D. Johnson¹ & James A. Huntington¹

Repeating intermolecular protein association by means of β -sheet expansion is the mechanism underlying a multitude of diseases including Alzheimer's, Huntington's and Parkinson's and the prion encephalopathies¹. A family of proteins, known as the serpins, also forms large stable multimers by ordered β -sheet linkages leading to intracellular accretion and disease². These 'serpinopathies' include early-onset dementia caused by mutations in neuroserpin, liver cirrhosis and emphysema caused by mutations in α_1 -antitrypsin (α_1 AT), and thrombosis caused by mutations in antithrombin³. Serpin structure and function are quite well understood, and the family has therefore become a model system for understanding the β -sheet expansion disorders collectively known as the conformational diseases⁴. To develop strategies to prevent and reverse these disorders, it is necessary to determine the structural basis of the intermolecular linkage and of the pathogenic monomeric state. Here we report the crystallographic structure of a stable serpin dimer which reveals a domain swap of more than 50 residues, including two long antiparallel β -strands inserting in the centre of the principal β -sheet of the neighbouring monomer. This structure explains the extreme stability of serpin polymers, the molecular basis of their rapid propagation, and provides critical new insights into the structural changes which initiate irreversible β -sheet expansion.

Serpin polymerization was first described for the Z variant of α_1 AT⁵. Liver biopsies demonstrated an accumulation of α_1 AT in the endoplasmic reticulum of hepatocytes from patients homozygous for the mutation, and electron micrographs of the purified inclusions showed a distinct beads-on-a-string morphology, with occasional self-terminating circular necklaces. Wild-type α_1 AT can be induced to form a similar beads-on-a-string morphology and corresponding ladders on native polyacrylamide gel electrophoresis (PAGE) by incubation under mild-denaturing conditions. The apparent repeated and ordered intermolecular linkage resulted in the phenomenon being called 'polymerization'⁶, and multiple mutations in several serpins have since been identified (indicated in Fig. 1a), most of which cause disease by the accumulation of polymers within cells⁷. As a protein family, serpins are particularly susceptible to the formation of stable polymers owing to the metastability of their native fold (Fig. 1a) and the thermodynamically driven β -sheet expansion mechanism required for their role as protease inhibitors^{8,9}. Approximately $-32 \text{ kcal mol}^{-1}$ is released by the incorporation of the reactive centre loop (RCL) of a serpin into the centre of its main β -sheet¹⁰ (sheet A, Fig. 1a and b). On the basis of the observations that serpin polymers had comparable hyperstability and that exogenous RCL peptides were capable of blocking polymerization, it was suggested that the polymeric linkage involved the incorporation of a portion of the RCL of one monomer into the β -sheet A of another⁶.

We set out to determine the molecular basis of serpin polymerization by protein crystallography. However, the inherent propensity of

polymers to continue to form higher order species after purification precluded concentration and crystallization. As mentioned previously, serpin polymers can self-terminate to form necklace structures, and a secreted self-terminating antithrombin dimer was recently observed in the blood plasma of patients carrying the P80S mutation¹¹. The dimer was composed of two hyperstable monomers and was incapable of continued polymerization, making this material amenable to crystallization. We were unable to obtain sufficient quantities of the dimeric mutant, so we produced and crystallized a self-terminating dimer from plasma-derived wild-type antithrombin (Methods and Supplementary Fig. 1), and solved its structure to 2.8 Å resolution (Supplementary Table and Supplementary Methods).

A single dimer was found in the asymmetric unit, composed of two latent-like (Fig. 1b) monomers arranged in a head-to-tail, back-to-back fashion (Fig. 1c). Close inspection of the electron density (Supplementary Fig. 2) revealed the presence of an extensive domain swap involving the insertion of strands 4 and 5 of one monomer into the β -sheet A of the other. The domain swap began immediately after helix I at residue 340 (Fig. 1c and Supplementary Fig. 2b), and continued to residue 390 at the end of strand 4A. Thus, residue Gly 339 is the first hinge point for a perfect domain swap that extends from the carboxy terminus of helix I to strands 5 and 4 of β -sheet A, with residues 340–390 occupying equivalent positions on the dimer partner as found in hyperstable monomeric conformers of antithrombin (latent or RCL-cleaved). From the end of strand 4A, the return to the parent monomer is direct, with no intermolecular or native contacts until residue 407 at the beginning of strand 5 of β -sheet B (Supplementary Fig. 2c). As with the latent conformer (Fig. 1b), residues C-terminal to the RCL—which form strand 1C in native antithrombin (400–405)—have been extracted from sheet C to provide sufficient length for return from strand 4A. The second hinge of the observed domain-swap is thus the loop between strands 1C and 5B, at residue 406. This large-scale domain swap completes β -sheet A in an unexpected fashion, and reveals the molecular basis behind the hyperstability of the polymeric linkage.

The structure of the self-terminating antithrombin dimer presented here fully explains the molecular pathology of the P80S variant. A question remains as to whether the observed intermolecular linkage involving strands 4A and 5A is specific for antithrombin or whether it is a general feature of serpin polymers. To investigate, we conducted the classic domain-swap experiment on the prototypical serpin α_1 AT. Cysteines were substituted at adjacent positions on strands 5 and 6A (indicated in Fig. 1b), so that intermolecular disulphide bonds should only form after polymerization when these strands are donated by different protomers. Although the inhibitory activity was unaffected, the oxidized monomeric form of the variant was fully resistant to polymerization (Fig. 2a), consistent with the requirement of strand 5A release for polymer formation. However, under reducing conditions the variant polymerized in a manner

¹University of Cambridge, Department of Haematology, Cambridge Institute for Medical Research, Wellcome Trust/MRC Building, Hills Road, Cambridge CB2 0XY, UK.

indistinguishable from wild type. Reoxidation of the polymeric material and subsequent denaturation by SDS showed the predicted intermolecular disulphide bonds (Fig. 2a), suggesting that the dual-strand domain swap (strands 4A/5A) observed for antithrombin also underlies the polymerization of α_1 AT.

For the practical reasons mentioned earlier, our crystallographic efforts were limited to the self-terminating dimer of antithrombin. Although this is sufficient to determine the intermolecular linkage, it was clear that the resulting structure would not represent the linear, self-propagating conformation, and that molecular modelling would be required. To create a dimer capable of continued polymerization, we linearized the self-terminating antithrombin dimer by breaking one intermolecular contact and placing the monomers in a head-to-tail

configuration (Fig. 3a). The resulting dimer was capable of self-propagation to form polymers of infinite length with no steric clashes (Fig. 3b). However, this configuration required the unravelling of residues within and just C-terminal to helix I, to create a flexible

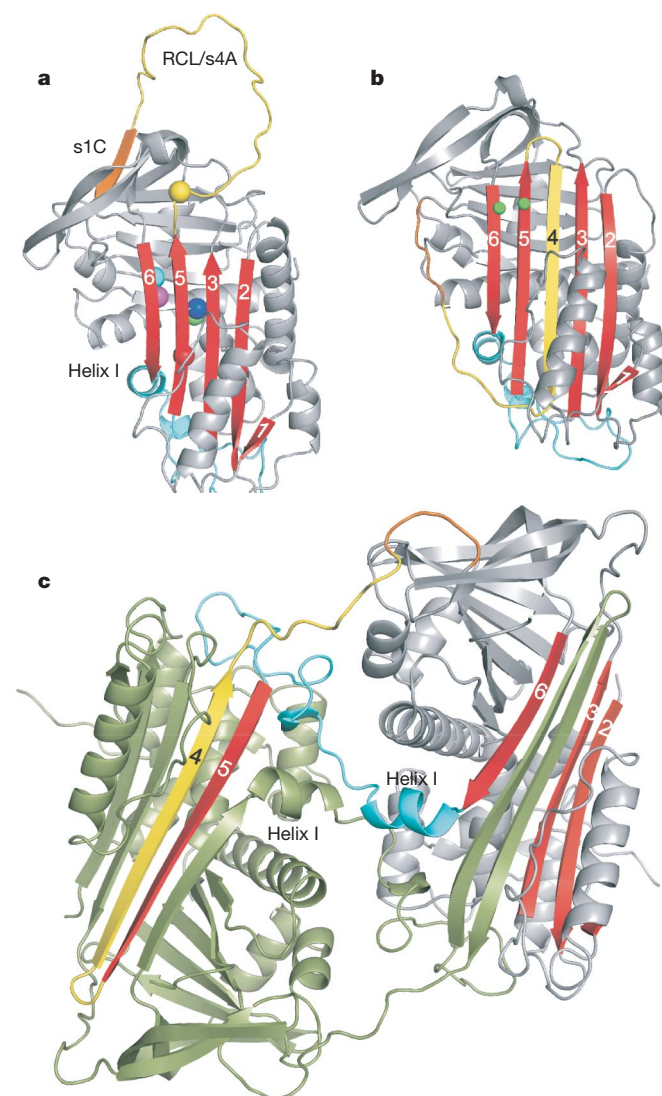


Figure 1 | Crystallographic structures of active, latent and self-terminating dimer of the serpin antithrombin. **a**, Active antithrombin is shown with the RCL on top (yellow) and the β -sheet A facing (red, with numbered strands). Either by proteolytic cleavage in the RCL or by extraction of strand 1C (s1C, orange), serpins incorporate the RCL into β -sheet A as strand 4A (s4A) resulting in a hyperstable six-stranded conformation. Polymerogenic mutations are shown (yellow is Z, magenta is Mmalton, blue is Siyama and Syracuse, red and cyan are His338Arg and Gly392Glu neuroserpin³, and green is P80S antithrombin¹¹). The loop connecting strand 6A to 5A is in cyan. **b**, The latent conformer of antithrombin is shown coloured as in **a**. Residues on strands 5 and 6A, which were mutated to Cys (see Fig. 2a), are indicated by green balls. **c**, The structure of the stable antithrombin dimer (monomer A coloured as in **b**, and monomer B is pale green).

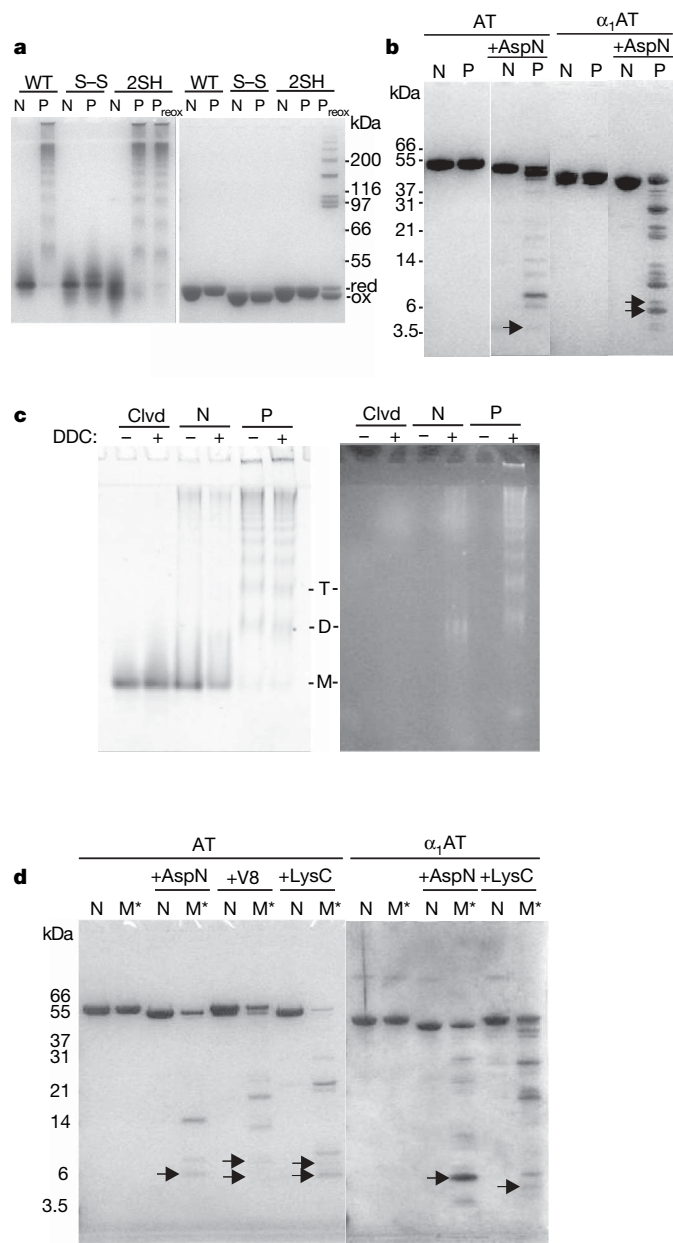


Figure 2 | Biochemical properties of serpin polymers and the M* state. **a**, The left panel is a native gel and the right panel is an SDS gel, and lanes are indicated as native (N), after polymerization (P), Cys-free control (WT), disulphide variant (S-S) and reduced variant (2SH). P_{reox} indicates chemical oxidation of the reduced variant polymers, and the positions of the oxidised (ox) and reduced (red) monomers are shown. Molecular mass standards are indicated. **b**, SDS-PAGE of AspN cleavage reaction for native (N) and polymerized (P) antithrombin (AT) and α_1 AT. Several of the bands were sequenced, and those indicated by the arrows are shown in Fig. 3a. **c**, Native PAGE visualized by Coomassie staining (left panel) or fluorescence (right panel) assesses the reactivity of a Cys residue in the centre of helix I in RCL-cleaved (clvd), native (N) and polymerized (P) α_1 AT. Lanes with '+' indicate that material was incubated with the fluorescent probe (didansylcysteine, DDC) just before electrophoresis. The positions of monomer (M), dimer (D) and trimer (T) are indicated. **d**, The proteolytic susceptibility of the native (N) and the M* state of antithrombin and α_1 AT was assessed using AspN, V8 and LysC by SDS-PAGE, as indicated. Arrows indicate cleavage sites shown in Fig. 3c.

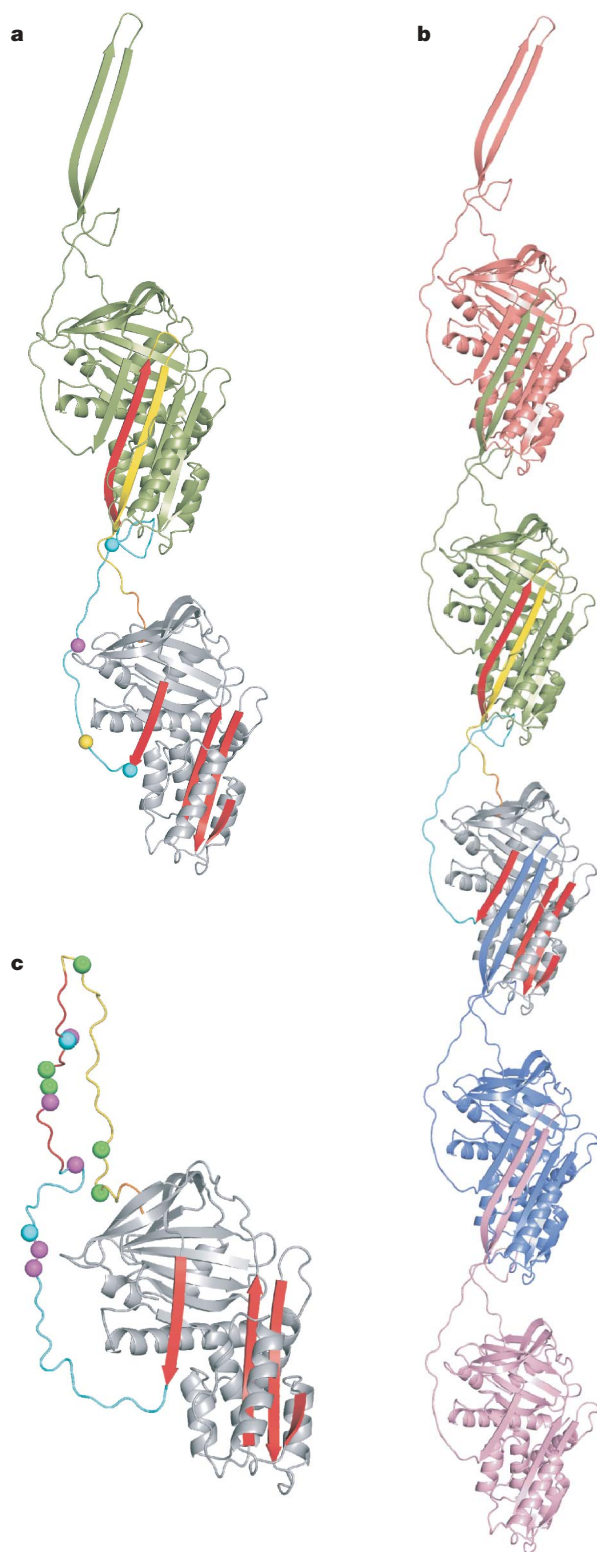


Figure 3 | Models of the serpin polymer and the M* state. **a**, A model of a linear serpin dimer (coloured as in Fig. 1c), with balls indicating AspN cleavage sites found in polymers of antithrombin (magenta) and α_1 AT (cyan). The yellow ball shows the position of the Leu303Cys substitution (Fig. 2c). **b**, A modelled pentamer was formed by the addition of protomers to both ends of the dimer. **c**, A model of the polymerogenic folding intermediate (M*) with the contiguous loop from the C terminus of strand 6A to strand 1C in a random-coil conformation. New proteolytic sites in this region (magenta balls for antithrombin, cyan balls for α_1 AT and green balls for PAI-1; ref. 20) were detected for the M* state (Fig. 2d).

intermolecular linker from a once stably folded region. To test this, we subjected native and polymeric antithrombin and α_1 AT to cleavage by the specific metalloprotease AspN. We found that polymers were consistently more susceptible to digestion by AspN than native controls (Fig. 2b), and amino-terminal sequencing revealed sites in and adjacent to helix I for polymers of both antithrombin and α_1 AT (plotted on Fig. 3a). The unravelling of helix I on polymerization of α_1 AT was assessed further by replacing a buried Leu in the centre of helix I with Cys (Leu303Cys position shown on Fig. 3a), and testing the reactivity of the sulphhydryl to fluorescent probes in monomeric and polymeric forms. The sulphhydryl was unreactive in native and RCL-cleaved states, consistent with a stable buried environment under helix I, but it was rapidly modified in the polymeric form (Fig. 2c), indicating the specific polymerization-induced unfolding of helix I.

Serpin polymerization *in vitro* is a two-step process, with an initial zero-order conversion to a polymorphic monomer (denoted as M*) followed by a slower concentration-dependent intermolecular association step¹². The M* state can be populated by high temperature, low pH or chaotropic agents, and slowly refolds to the native conformation when returned to normal conditions^{13–16}. In contrast, serpin polymers do not require special conditions to propagate, and isolated dimers, trimers and so on, rapidly convert to higher order polymers under native conditions¹⁷. This property, recently referred to as ‘infectivity’, requires active donor and acceptor ends to persist when conditions are returned to those favouring the native state. Our structure provides a ready explanation for polymer infectivity. As illustrated in Fig. 3a, the top monomer of the linear dimer is compelled to remain in a state capable of donating strands 4 and 5A because its own β -sheet A is already complete, whereas the bottom monomer must remain in a quiescent acceptor conformation because its own strands 4 and 5 are inserted into the upper monomer. The model also explains why polymers are capable of incorporating exogenous peptides (possibly into the vacated strand 5A position of the bottom monomer), but not the RCLs (strand 4A) of native or peptide-complexed monomeric forms of the same serpin¹⁷ (because strand 5A is also required).

To understand pathological serpin polymerization, why certain mutations predispose serpins to polymerize, and to develop strategies to prevent or slow polymerization, it is necessary to define the structural features of the M* state. One important implication of our structure is that the principal feature of M* is the exposure of strand 5A. The M* state should thus resemble the monomeric component of the linear dimer (Fig. 3c), with the contiguous loop from the C terminus of strand 6A to the N terminus of strand 5B in an exposed and flexible conformation. This hypothesis is supported by the slower migration of M* on native PAGE¹⁸ and unfolding studies which showed that strand 1C and the loop between strand 5A and helix I are two of the first elements to unfold, and predictably, two of the last to fold¹⁹. To investigate the predicted unfolding of this region in the M* state, we subjected antithrombin and α_1 AT to limited proteolysis after brief exposure to conditions known to populate M* (Fig. 2d). The M* state was generally more susceptible to proteolysis than the native controls, and sequencing of the unique bands demonstrated cleavage sites in strand 5A and C-terminal to helix I for both serpins. Similar cleavage sites have been described for another serpin (PAI-1) after incubation with small hydrophobic compounds²⁰ (all sites plotted on Fig. 3c). The location of the pathological serpin mutations, on and underlying strand 5A, can now be understood in terms of a slowing of the final folding step from a polymorphic intermediate with an exposed strand 5A to the native state with the strand incorporated in β -sheet A^{13,18,21}.

Domain swapping is recognized as a general mechanism behind certain conformational diseases²², and although there are several examples in the literature of domain-swapped oligomers²³, none represents the pathological linkage that results in a protein deposition disease. Our crystal structure is unique in this sense, and also differs in using two hinges and an internal motif, and in the hyperstability

of the intermolecular linkage. Our data suggests that the free-energy gained by completing β -sheet A is used to unfold helix I and the following coiled region. This unfolding event exposes a 30-residue linker that might explain why serpin polymers are hydrophobic²⁴, prone to precipitation^{25,26} and sensitive to seeding²⁷. The linker is ~50% hydrophobic and is predicted to have a high β -strand propensity. Therefore, the linker could be responsible for the lateral association of linear serpin polymers into tangled aggregates, like those observed by electron microscopy from intact Collins bodies of neuroserpin²⁸. The precise mechanism of lateral association is unclear, but it is probable that the molecular pathology of serpin polymerization depends on such a secondary event for deposition and evasion of the cellular housekeeping machinery. Our findings have a profound effect on efforts to prevent, slow or reverse serpin polymerization. Strategies might target the M* state by increasing its rate of folding or by blocking the exposed strand 5A site. Alternatively, it may be possible to target the secondary lateral association event with peptidic β -sheet breakers²⁹.

METHODS SUMMARY

The stable antithrombin dimer was produced from plasma-derived native protein and purified by gel filtration. Crystals were grown by vapour diffusion. Data processing and refinement statistics are given in the Supplementary Table.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 30 April; accepted 2 September 2008.

Published online 15 October 2008.

- Harrison, R. S., Sharpe, P. C., Singh, Y. & Fairlie, D. P. Amyloid peptides and proteins in review. *Rev. Physiol. Biochem. Pharmacol.* **159**, 1–77 (2007).
- Lomas, D. A. & Carrell, R. W. Serpinopathies and the conformational dementias. *Nature Rev. Genet.* **3**, 759–768 (2002).
- Lomas, D. A. *et al.* Molecular mousetraps and the serpinopathies. *Biochem. Soc. Trans.* **33**, 321–330 (2005).
- Carrell, R. W. & Lomas, D. A. Conformational disease. *Lancet* **350**, 134–138 (1997).
- Sharp, H. L., Bridges, R. A., Krivit, W. & Freier, E. F. Cirrhosis associated with α_1 -antitrypsin deficiency: a previously unrecognized inherited disorder. *J. Lab. Clin. Med.* **73**, 934–939 (1969).
- Lomas, D. A., Evans, D. L., Finch, J. T. & Carrell, R. W. The mechanism of Z α_1 -antitrypsin accumulation in the liver. *Nature* **357**, 605–607 (1992).
- Lomas, D. A. *et al.* Polymerisation underlies α_1 -antitrypsin deficiency, dementia and other serpinopathies. *Front. Biosci.* **9**, 2873–2891 (2004).
- Huntington, J. A., Read, R. J. & Carrell, R. W. Structure of a serpin-protease complex shows inhibition by deformation. *Nature* **407**, 923–926 (2000).
- Gettins, P. G. Serpin structure, mechanism, and function. *Chem. Rev.* **102**, 4751–4804 (2002).
- Im, H., Ahn, H. Y. & Yu, M. H. Bypassing the kinetic trap of serpin protein folding by loop extension. *Protein Sci.* **9**, 1497–1502 (2000).
- Corral, J. *et al.* Mutations in the shutter region of antithrombin result in formation of disulfide-linked dimers and severe venous thrombosis. *J. Thromb. Haemost.* **2**, 931–939 (2004).
- Dafforn, T. R., Mahadeva, R., Elliott, P. R., Sivasothy, P. & Lomas, D. A. A kinetic mechanism for the polymerization of α_1 -antitrypsin. *J. Biol. Chem.* **274**, 9548–9555 (1999).
- Kim, D. & Yu, M. H. Folding pathway of human α_1 -antitrypsin: characterization of an intermediate that is active but prone to aggregation. *Biochem. Biophys. Res. Commun.* **226**, 378–384 (1996).
- Devlin, G. L., Chow, M. K., Howlett, G. J. & Bottomley, S. P. Acid denaturation of α_1 -antitrypsin: characterization of a novel mechanism of serpin polymerization. *J. Mol. Biol.* **324**, 859–870 (2002).
- Tew, D. J. & Bottomley, S. P. Probing the equilibrium denaturation of the serpin α_1 -antitrypsin with single tryptophan mutants: evidence for structure in the urea unfolded state. *J. Mol. Biol.* **313**, 1161–1169 (2001).
- Egelund, R. *et al.* A regulatory hydrophobic area in the flexible joint region of plasminogen activator inhibitor-1, defined with fluorescent activity-neutralizing ligands. Ligand-induced serpin polymerization. *J. Biol. Chem.* **276**, 13077–13086 (2001).
- Zhou, A. & Carrell, R. W. Dimers initiate and propagate serine protease inhibitor polymerisation. *J. Mol. Biol.* **375**, 36–42 (2008).
- Yu, M. H., Lee, K. N. & Kim, J. The Z type variation of human α_1 -antitrypsin causes a protein folding defect. *Nature Struct. Biol.* **2**, 363–367 (1995).
- James, E. L., Whisstock, J. C., Gore, M. G. & Bottomley, S. P. Probing the unfolding pathway of α_1 -antitrypsin. *J. Biol. Chem.* **274**, 9482–9488 (1999).
- Kjoller, L. *et al.* Conformational changes of the reactive-centre loop and β -strand 5A accompany temperature-dependent inhibitor-substrate transition of plasminogen-activator inhibitor 1. *Eur. J. Biochem.* **241**, 38–46 (1996).
- Jung, C. H., Na, Y. R. & Im, H. Retarded protein folding of deficient human α_1 -antitrypsin D256V and L41P variants. *Protein Sci.* **13**, 694–702 (2004).
- Bennett, M. J., Sawaya, M. R. & Eisenberg, D. Deposition diseases and 3D domain swapping. *Structure* **14**, 811–824 (2006).
- Liu, Y. & Eisenberg, D. 3D domain swapping: as domains continue to swap. *Protein Sci.* **11**, 1285–1299 (2002).
- James, E. L. & Bottomley, S. P. The mechanism of α_1 -antitrypsin polymerization probed by fluorescence spectroscopy. *Arch. Biochem. Biophys.* **356**, 296–300 (1998).
- Pearce, M. C., Cabrita, L. D., Ellisdon, A. M. & Bottomley, S. P. The loss of tryptophan 194 in antichymotrypsin lowers the kinetic barrier to misfolding. *FEBS J.* **274**, 3622–3632 (2007).
- Pedersen, K. E. *et al.* Plasminogen activator inhibitor-1 polymers, induced by inactivating amphipathic organochemical ligands. *Biochem. J.* **372**, 747–755 (2003).
- Crowther, D. C., Serpell, L. C., Dafforn, T. R., Gooptu, B. & Lomas, D. A. Nucleation of α_1 -antichymotrypsin polymerization. *Biochemistry* **42**, 2355–2363 (2003).
- Davis, R. L. *et al.* Familial dementia caused by polymerization of mutant neuroserpin. *Nature* **401**, 376–379 (1999).
- Soto, C. *et al.* β -sheet breaker peptides inhibit fibrillogenesis in a rat brain model of amyloidosis: implications for Alzheimer's therapy. *Nature Med.* **4**, 822–826 (1998).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank R. Carrell, D. Crowther and D. Lomas for their comments on the manuscript. We are also grateful to J. Löwe for access to the robotic crystallisation facility in the MRC-LMB, and to M. Weldon for N-terminal sequencing. This work was supported by the National Institutes of Health (USA) and the Uehara Memorial Foundation (Japan, to M.Y.), and J.A.H. is a senior MRC non-clinical fellow. Data were collected at beamlines I02 and I04 at the Diamond Light Source and we acknowledge the support of L. Duke, G. Evans, R. Flaig, J. Sandy and T. Sorensen.

Author Information Atomic coordinates and structure factors have been deposited in the Protein Data Bank under the accession code 2ZNN. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to J.A.H. (jah52@cam.ac.uk).

METHODS

Dimer preparation. The native α -glycoform of antithrombin was purified from outdated fresh-frozen human blood plasma, as previously described³⁰. The stable dimer was produced by incubating 0.5 mg ml⁻¹ antithrombin in 10 mM sodium acetate, pH 5.7, at 37 °C for 16 h. Supplementary Fig. 1a shows a typical time course under these conditions and the accumulation of the stable dimer. It migrated slightly faster than the polymerization-competent dimer on native PAGE, suggesting a more compact structure. Thrombin was unable to cleave the stable dimer, whereas precisely one monomer from the polymerization-competent dimer was cleaved in the RCL (Supplementary Fig. 1b). These data support the conclusion that the stable dimer was self-terminating by intermolecular RCL- β -sheet-A interactions. The stable dimer could also be formed by incubation at pH 7 and increasing the temperature to 53 °C, although the yields were low. The 16 h incubation ensured that no native antithrombin remained and that the stable dimer was the only dimeric species. The reaction mixture was concentrated, buffer exchanged into 50 mM Tris, pH 7.4, 0.3 M NaCl, and loaded onto and eluted from a Superdex 200 HR10/30 size exclusion column (GE Healthcare). Fractions containing the dimer were pooled, concentrated to 5 mg ml⁻¹ in 10 mM Tris, pH 7.4, 50 mM NaCl for crystallization trials.

Crystallization, data collection, molecular replacement and refinement. Initial hits were obtained using the robotic crystallization facility in the MRC-LMB³¹ from a screen of 384 conditions in 96-well sitting-drop vapour diffusion plates. Optimisation was conducted by hand in hanging-drop VDX plates, and diffraction quality crystals were eventually obtained in 20% PEG 3350, 0.2 M tripotassium citrate. An initial, low-resolution data set was obtained from four crystals flash cooled in paratone-N on Diamond Light Source beamline I02. The data were processed using Mosflm, Scala and Truncate³², and were of sufficient quality ($R_{\text{merge}} = 0.246$, mean $(I/\sigma I) = 7.0$) and resolution (3.2 Å) for molecular replacement and refinement. Molecular replacement solutions were found for two copies of antithrombin using native and latent antithrombin conformers (1E04) with the program Phaser³³; the latent conformer gave the best score. Examination of initial electron density (calculated from molecular replacement solutions using either native or latent antithrombin) showed that each monomer possessed a fully six-stranded β -sheet A with unbroken density for the turn between strands 5 and 4A. The structure was refined using heavy non-crystallographic symmetry (NCS) restraints for a large subset of the residues with the program CNS³⁴ (version 1.0) and rebuilt using XtalView³⁵ and Coot³⁶. The domain swap was evident after the first round of refinement, and the linker regions were built into unequivocal positive difference density. We subsequently obtained a higher resolution data set from a single crystal on beamline I04 (wavelength of 0.919 Å) after air dehydration and cryoprotection in paratone-N (100K). The R_{free} set was taken from the initial data set and extended randomly to 2.8 Å resolution, and the low-resolution model was subjected to rigid body refinement, with each core monomer constituting a rigid body. Heavy NCS restraints were used initially and subsequently relaxed, until in the final rounds NCS restraints were dropped altogether. Data processing and refinement statistics for the high-resolution data and structure are given in the Supplementary Table. The RCSB validation server placed 98.5% of the residues in the most favoured and additionally allowed Ramachandran regions, with no residues in the disallowed region. A similar calculation on the MolProbity server³⁷ classed 92.8% as favoured and 1.1% as outliers. Modelling an open α_1 AT dimer was conducted using the coordinates of RCL cleaved α_1 AT (PDB accession 1EZK) and programs XtalView³⁵ and CNS³⁴. Figures were made using PyMol³⁸.

α_1 AT Cys variants. The α_1 AT used in this study was produced recombinantly from *Escherichia coli* as reported previously³⁹. The template α_1 AT was a double variant with a reactive centre mutation Met \rightarrow Arg and the only Cys residue (232) mutated to Ala. Mutagenesis was conducted using the Quick Change kit and associated protocol (Stratagene). This material was used for the limited proteolysis studies and was the background for mutations in which specific Cys residues were introduced. To test whether strand 5A is part of the domain swap for α_1 AT, we introduced Cys residues on adjacent, externally exposed positions on strands 5A and 6A (Thr339Cys and Ser292Cys, respectively). The disulphide bond spontaneously formed during purification of the variant, as indicated by faster mobility on SDS-PAGE. The inhibitory activity of the variant was indistinguishable from wild type, indicating that β -sheet A expansion was unperturbed by the intramolecular disulphide bond between strands 5 and 6A. After reduction of the disulphide bond (15 min incubation in 2 mM dithiothreitol (DTT) at 37 °C), polymerization was induced by incubation in 0.75 M guanidine-HCl for 8 h at 37 °C in 50 mM sodium phosphate buffer, pH 7.0. The reaction mixture was

subsequently dialysed against Tris-buffered saline overnight at 4 °C, and the disulphide bond was re-formed by adding tetramethylazodicarboamide to 0.5 mM for 2 h at room temperature. Wild type and the oxidized variant were treated identically, but without the initial incubation with DTT. Samples were run on 8% native and 10% SDS-PAGE. To test the unfolding of helix I in the α_1 AT polymers, Leu 303 was substituted by Cys, and the accessibility of the sulphhydryl was probed for native, RCL-cleaved and polymers by reacting with DDC (Sigma). The inhibitory activity of the Lys303Cys variant was indistinguishable from wild type, but was mildly polymerogenic. The three forms of the variant were incubated with 0.8 mM DDC for 5 min at room temperature in the dark, just before loading on native PAGE. Similar results were obtained using the probe 5-iodoacetamidofluorescein (data not shown).

Limited proteolysis. The exposure of distinct regions on polymerization of antithrombin and α_1 AT was tested by limited proteolysis using the highly specific metalloprotease, AspN, which cleaves peptide bonds N-terminal to available Asp residues. Polymers of antithrombin were prepared by incubating 0.5 mg ml⁻¹ antithrombin in 10 mM sodium phosphate buffer (pH 6.5) at 53 °C for 1 to 2 h. Native and polymeric forms were then subjected to AspN digestion (1:65 ratio by weight) overnight at 4 °C. Reactions were stopped by adding SDS-loading buffer (including DTT) and boiling for 5 min. The samples were then run on a 12% Bis-Tris gel in MES buffer (Invitrogen). To identify the bands of interest, identical samples were run on a separate gel and blotted onto a PVDF membrane for N-terminal sequencing. All N-terminal sequencing was conducted by the Protein and Nucleic Acid Chemistry Facility, Department of Biochemistry, University of Cambridge. Polymers of recombinant α_1 AT were generated by incubation of 0.025 mg ml⁻¹ α_1 AT in 50 mM sodium phosphate buffer, pH 7.0, at 53 °C for 3 h. Native and polymeric forms were subject to AspN digestion (1:50 ratio by weight) overnight at 4 °C. Samples were then precipitated with 10% TCA and washed with acetone. Pellets were resuspended in 50 mM Tris, pH 8.4, and subjected to SDS-PAGE and N-terminal sequencing as above. Identical results were obtained when polymers of α_1 AT were generated with 0.75 M guanidine-HCl at 37 °C. The M* form of antithrombin was generated by incubation of 0.25 mg ml⁻¹ antithrombin in 10 mM sodium citrate buffer (pH 4.5) at room temperature for 4 h, followed by rapid reformulation into 50 mM Tris, 100 mM NaCl, pH 7.4, and the addition of proteases AspN, V8 or LysC (1:50–1:100 ratio by weight). After an overnight incubation at 4 °C, the reactions were stopped by the addition of SDS-loading buffer and boiling. The samples were then subjected to SDS-PAGE and N-terminal sequencing as above. The M* form of α_1 AT was generated by incubating 0.1 mg ml⁻¹ of α_1 AT in 50 mM sodium phosphate buffer, pH 7.0, containing 0.8 M guanidine-HCl at 37 °C for 2 h. Native and M* (after reformulation as for antithrombin) α_1 AT were subject to AspN and LysC digestion as above (V8 cleaves α_1 AT in the RCL and was therefore not used). Samples were then precipitated with 10% TCA and washed with acetone. Pellets were resuspended in 50 mM Tris, pH 8.4, and subjected to SDS-PAGE and N-terminal sequencing as above. Similar cleavage results were obtained for α_1 AT when M* was generated by short incubations either at low pH (pH 4.5 and 37 °C) or at high temperature (pH 7 and 49 °C).

30. Olson, S. T., Bjork, I. & Shore, J. D. Kinetic characterization of heparin-catalyzed and uncatalyzed inhibition of blood coagulation proteinases by antithrombin. *Methods Enzymol.* **222**, 525–559 (1993).
31. Stock, D., Perisic, O. & Lowe, J. Robotic nanolitre protein crystallisation at the MRC Laboratory of Molecular Biology. *Prog. Biophys. Mol. Biol.* **88**, 311–327 (2005).
32. Leslie, A. W. G. in *Joint CCP4 and ESF-EACMB Newsletter on Protein Crystallography* (Daresbury Laboratory, 1992).
33. McCoy, A. J., Grosse-Kunstleve, R. W., Storoni, L. C. & Read, R. J. Likelihood-enhanced fast translation functions. *Acta Crystallogr. D* **61**, 458–464 (2005).
34. Brunger, A. T. et al. Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
35. McRee, D. E. A visual protein crystallographic software system for X11/XView. *J. Mol. Graph.* **10**, 44–46 (1992).
36. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
37. Lovell, S. C. et al. Structure validation by ϕ , ψ and C β deviation. *Proteins* **50**, 437–450 (2003).
38. DeLano, W. L. The PyMOL Molecular Graphics System. <<http://www.pymol.org>> (2002).
39. Zhou, A., Carrell, R. W. & Huntington, J. A. The serpin inhibitory mechanism is critically dependent on the length of the reactive center loop. *J. Biol. Chem.* **276**, 27541–27547 (2001).

ROS3 is an RNA-binding protein required for DNA demethylation in *Arabidopsis*

Xianwu Zheng¹, Olga Pontes², Jianhua Zhu³, Daisuke Miki¹, Fei Zhang^{1,4}, Wen-Xue Li¹, Kei Iida¹, Avnish Kapoor¹, Craig S. Pikaard² & Jian-Kang Zhu¹

DNA methylation is an important epigenetic mark for transcriptional gene silencing (TGS) in diverse organisms^{1–6}. Recent studies suggest that the methylation status of a number of genes is dynamically regulated by methylation and demethylation^{7–10}. In *Arabidopsis*, active DNA demethylation is mediated by the ROS1 (repressor of silencing 1) subfamily of 5-methylcytosine DNA glycosylases through a base excision repair pathway^{8,10–13}. These demethylases have critical roles in erasing DNA methylation and preventing TGS of target genes^{7,8,10}. However, it is not known how the demethylases are targeted to specific sequences. Here we report the identification of ROS3, an essential regulator of DNA demethylation that contains an RNA recognition motif. Analysis of *ros3* mutants and *ros1 ros3* double mutants suggests that ROS3 acts in the same genetic pathway as ROS1 to prevent DNA hypermethylation and TGS. Gel mobility shift assays and analysis of ROS3 immunoprecipitate from plant extracts shows that ROS3 binds to small RNAs *in vitro* and *in vivo*. Immunostaining shows that ROS3 and ROS1 proteins co-localize in discrete foci dispersed throughout the nucleus. These results demonstrate a critical role for ROS3 in preventing DNA hypermethylation and suggest that DNA demethylation by ROS1 may be guided by RNAs bound to ROS3.

We developed a sensitive assay system in *Arabidopsis* to dissect active DNA demethylation genetically^{10,14}. The system consists of the *RD29A-LUC* transgene (firefly luciferase reporter driven by the stress-responsive *RD29A* promoter) and the non-allelic endogenous *RD29A* gene. The *RD29A* promoter is subjected to continuous short interfering RNA (siRNA)-directed DNA methylation such that active DNA demethylation is required to keep the *RD29A* and *RD29A-LUC* genes transcriptionally active. In *ros1* mutants, the *RD29A* promoter for both the transgene and endogenous gene becomes hypermethylated and both genes are silenced¹⁰. In addition, the *35S-NPTII* transgene linked to *RD29A-LUC* is also silenced such that *ros1* mutant plants are sensitive to kanamycin. We isolated the *ros3* mutant from a T-DNA mutagenized population¹⁵. Like *ros1* mutations, *ros3* causes a substantial reduction in bioluminescence emission (Fig. 1a and Supplementary Fig. 1a, b) as well as sensitivity to kanamycin (Fig. 1b). Genetic analysis indicated that the *ros3* mutation is recessive and affects a nuclear gene (data not shown).

Northern blot (Fig. 1c) and nuclear run-on (Fig. 1d) assays showed that the *RD29A-LUC* and *35S-NPTII* transgenes and the endogenous *RD29A* gene are repressed transcriptionally in *ros3* plants. Compared with the wild type, both the endogenous (Fig. 2a) and transgene (Fig. 2b) *RD29A* promoters in *ros3* mutants are substantially more heavily methylated at CpG, CpNpG (where N is A, T or C) and CpNpN sites (Supplementary Table 1). Southern blot analysis with methylation-sensitive restriction enzymes also indicated DNA

hypermethylation in the *RD29A* promoters in the mutant (Supplementary Fig. 2a, b). Treatment with the cytosine methylation inhibitor 5-aza-2'-deoxycytidine increased *RD29A-LUC* expression in the *ros3* mutant to wild-type levels (Supplementary Fig. 3). These results suggest that DNA hypermethylation is responsible for the TGS in *ros3* mutant plants.

The *nrpd1a-1* mutation in the largest subunit of RNA polymerase IVa blocks the accumulation of 24-nucleotide siRNAs corresponding to the *RD29A* promoter (data not shown). Analysis of *nrpd1a ros3* double mutants showed that the *nrpd1a* mutation causes a significant increase in *RD29A-LUC* expression (Fig. 2c, d) and a substantial decrease in CpG, CpNpG and CpNpN methylation at the transgene *RD29A* promoter (Fig. 2e) in the *ros3* mutant background. The release of TGS in *ros3* by *nrpd1a* suggests that the siRNAs targeting the *RD29A* promoter are the initial trigger of TGS in the *ros3* mutant. The levels of siRNAs targeting the *RD29A* promoter do not differ substantially between *ros3* and the wild type (Supplementary Fig. 1c).

To analyse the genetic interaction between *ros3* and *ros1*, we crossed *ros1-1* with *ros3* and generated a *ros1 ros3* double mutant. Analysis of cytosine methylation levels at the endogenous *RD29A* as well as transgenic *RD29A* promoters showed that the levels of CpG, CpNpG and asymmetric CpNpN methylation in *ros1 ros3* are similar to those in the *ros1* single mutant (Fig. 2a, b and Supplementary Table 1). DNA methylation at the intergenic 80, At1g23740389–0769, At4g5390602–0993 and At2g1699280–9430 loci is higher in *ros1* and *ros3* single mutants compared to the wild type, and the double mutant shows a methylation pattern similar to *ros1* (Supplementary Fig. 4a–d and Supplementary Table 2). Therefore, the effects of the *ros1* and *ros3* mutations on the methylation levels are not additive, indicating that ROS3 and ROS1 function in the same genetic pathway to demethylate DNA.

The second and third pairs of true leaves in *ros3* are narrower and slightly lobed compared to those in the wild type (Supplementary Fig. 5a). On the basis of the leaf morphology and kanamycin sensitivity phenotypes, we mapped the *ros3* mutation to a genomic region on chromosome V (Supplementary Fig. 5b). Sequencing of all annotated genes in this region revealed a deletion of 260 base pairs (bp) in *ros3* that includes the first intron and part of the second exon of AT5G58130 (Supplementary Figs 5c and 6). AT5G58130 encodes a protein with an amino-terminal RNA recognition motif (RRM), a central COG5406 domain found in the nucleosome binding factor SPN, and a carboxy-terminal sequence with a secondary structure similar to those of RRMs (Supplementary Figs 5c, 7 and 8). A wild-type genomic fragment of AT5G58130 was found to complement the luminescence, leaf morphology (Supplementary Fig. 5d, e) and DNA hypermethylation (Fig. 2a, b) phenotypes of the *ros3* mutant, confirming that AT5G58130 is ROS3.

¹Center for Plant Cell Biology and Department of Botany and Plant Sciences, University of California, Riverside, California 92521, USA. ²Biology Department, Washington University, Campus Box 1137, One Brookings Drive, St Louis, Missouri 63130, USA. ³Department of Plant Science and Landscape Architecture, University of Maryland, College Park, Maryland 20742, USA. ⁴Plant Biotechnology Research Center, School of Agriculture and Biology, Shanghai Jiao Tong University, Shanghai 200240, China.

Expression of a translational fusion of yellow fluorescence protein (YFP) at the N terminus of ROS3 in onion epidermal cells or *Arabidopsis* showed a clear nuclear localization of the fusion protein (Supplementary Fig. 9a). Close inspection revealed that the fusion protein is concentrated in discrete foci in the nucleoplasm and is also present in the nucleolus (Supplementary Fig. 9b, c). Analysis of *GUS* reporter gene expression driven by the *ROS3* promoter suggested that *ROS3* is ubiquitously expressed in plant tissues (Supplementary Fig. 9d).

Because *ROS3* contains an RRM domain, we tested whether it may bind to RNA. We expressed recombinant *ROS3* or its truncated forms with an N-terminal histidine tag (Fig. 3a, b). Electrophoretic mobility-shift assays with single-stranded and double-stranded RNAs of different sizes (Supplementary Table 3) were carried out. *ROS3* was able to bind single-stranded RNA of sequence b (corresponding to the *RD29A* promoter) that was either 21-, 24- or 26-nucleotides, but it did not bind 21-, 24- or 26-nucleotide single-stranded RNA of

sequence a (Fig. 3c). It also did not bind a 40-nucleotide single-stranded RNA (Fig. 3c). There was no binding (data not shown) to any double-stranded RNA or any other single-stranded RNA tested (Supplementary Table 3). Removal of the N-terminal RRM domain and the C-terminal region (Fig. 3c, lanes 4, 12 and 14; Fig. 3d, lane 4), or of the N-terminal RRM domain alone (Fig. 3d, lane 2), abolished binding to RNA of sequence b. However, removal of the C-terminal region did not substantially affect RNA binding (Fig. 3d, lane 3). The binding to the 24-nucleotide RNA (sequence b) is dependent on *ROS3* protein concentration, and is competed by excess unlabelled small RNA of the same sequence (Fig. 3e). These results show that *ROS3* has the capacity to bind specific small RNAs and that the RRM domain is required for this binding. Although *ROS3* could bind 21-, 24- and 26-nucleotide single-stranded RNAs of sequence b *in vitro*, RNA blot analysis showed that only a ~24-nucleotide small RNA of sequence b is present in *Arabidopsis* (Supplementary Fig. 10a).

To determine whether *ROS3* may bind small RNAs *in vivo*, *ROS3* protein was immunoprecipitated from *Arabidopsis* extracts using anti-*ROS3* antisera. The immunoprecipitate was fractionated on a

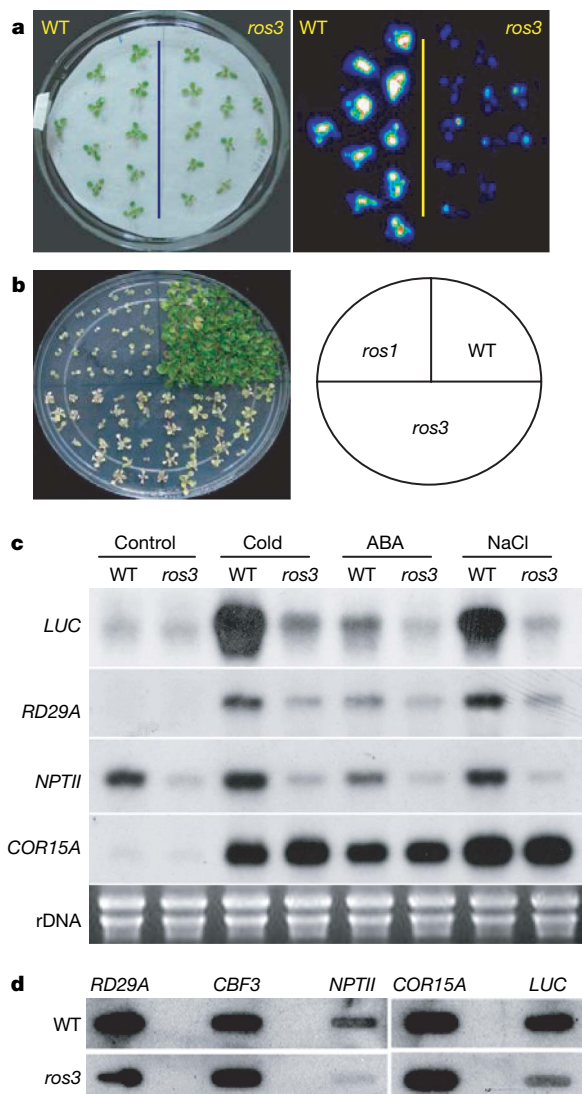


Figure 1 | The *ros3* mutation causes transcriptional gene silencing. **a**, Stress-induced expression of the *RD29A-LUC* transgene in wild type (WT) and *ros3* mutant plants after treatment with 300 mM NaCl for 5 h. **b**, Like *ros1*, the *ros3* mutant plants are sensitive to kanamycin. **c**, Northern blots showing that *ros3* reduces the transcript levels of *LUC*, *NPTII* and endogenous *RD29A*, but not of the control, *COR15A*. Plants were either untreated (control) or treated with cold (4 °C) for 24 h, 100 μM abscisic acid (ABA) for 3 h, or 300 mM NaCl for 5 h. **d**, Nuclear run-on assay showing the pre-mRNA levels of *LUC*, *NPTII* and *RD29A* genes in wild-type and *ros3*. *COR15A* and *CBF3* were used as controls.

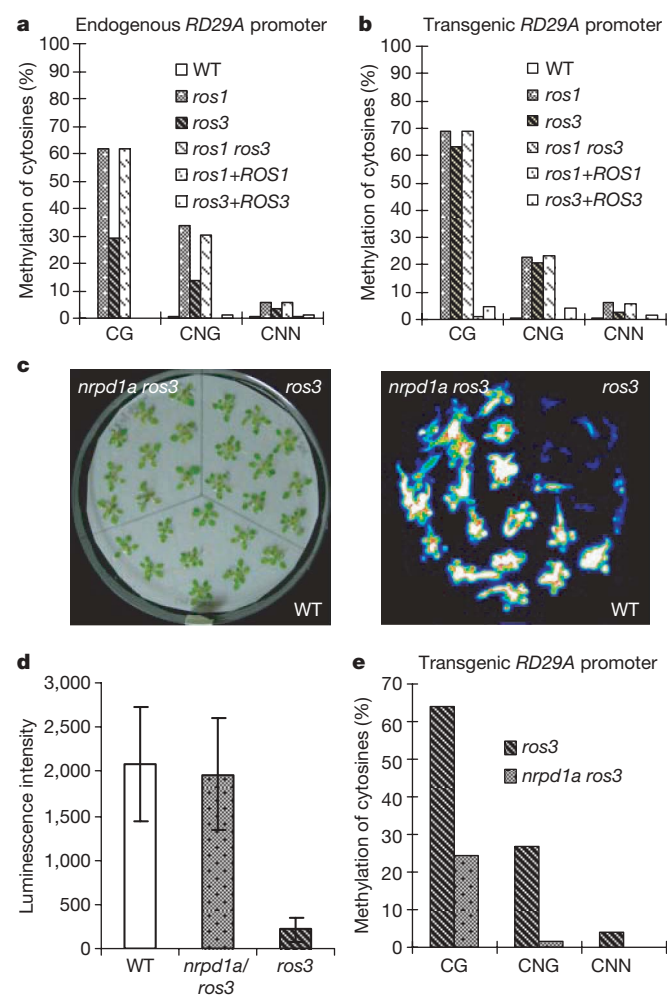


Figure 2 | DNA hypermethylation in *ros3* and suppression of *ros3* by *nrpd1a*. **a**, **b**, Bisulphite sequencing analysis of promoter methylation status of the endogenous *RD29A* (**a**) and *RD29A-LUC* transgene (**b**) in wild type, *ros1*, *ros3*, *ros1 ros3*, *ros1* complemented with wild-type *ROS1* transgene, and *ros3* complemented with wild-type *ROS3* transgene. **c**, Suppression of *ros3* by the *nrpd1a-1* mutation. Seedlings of wild type, *ros3* and *nrpd1a ros3* double mutants grown in MS medium for 3 weeks were transferred to a filter paper soaked with 300 mM NaCl for 5 h. Left, picture of the seedlings; right, luminescence image. **d**, Quantification of luminescence in **c**. Numbers on the y axis represent counts per seedling. Error bars represent standard deviation ($n = 20$). **e**, DNA methylation status at the transgene *RD29A* promoter in *ros3* and *nrpd1a ros3* plants.

polyacrylamide gel and putative ROS3-bound small RNAs of ~15–30 nucleotides were extracted from the gel, cloned and sequenced. Out of 288 clones sequenced, 140 had inserts of 10–30 nucleotides (Supplementary Table 4). Notably, the sequences are very rich in Gs. Two 25-nucleotide sequences—RNA sequence c, 5'-GGGAG UCCGGAGACGUCGGCGGGG-3' and RNA sequence d, 5'-UCGGGAGGGAAGCGGAUGGGGGCCG-3'—were chosen for further analysis. These small RNAs correspond to a genomic region (intergenic between At3G41979 and At3TE58310) that appears to have higher DNA methylation in the *ros1 dml2 dml3* demethylase triple mutant than in Columbia wild-type control plants⁹ (<http://neomorph.salk.edu/epigenome.html>). Both RNA sequence c and RNA sequence d could be detected as ~25-nucleotide small RNAs in plants by northern blot analysis (Supplementary Fig. 10b, c). ROS3 protein is able to bind RNA sequence c and RNA sequence d *in vitro* (Fig. 3f). Bisulphite sequencing analysis indicated higher levels of CG,

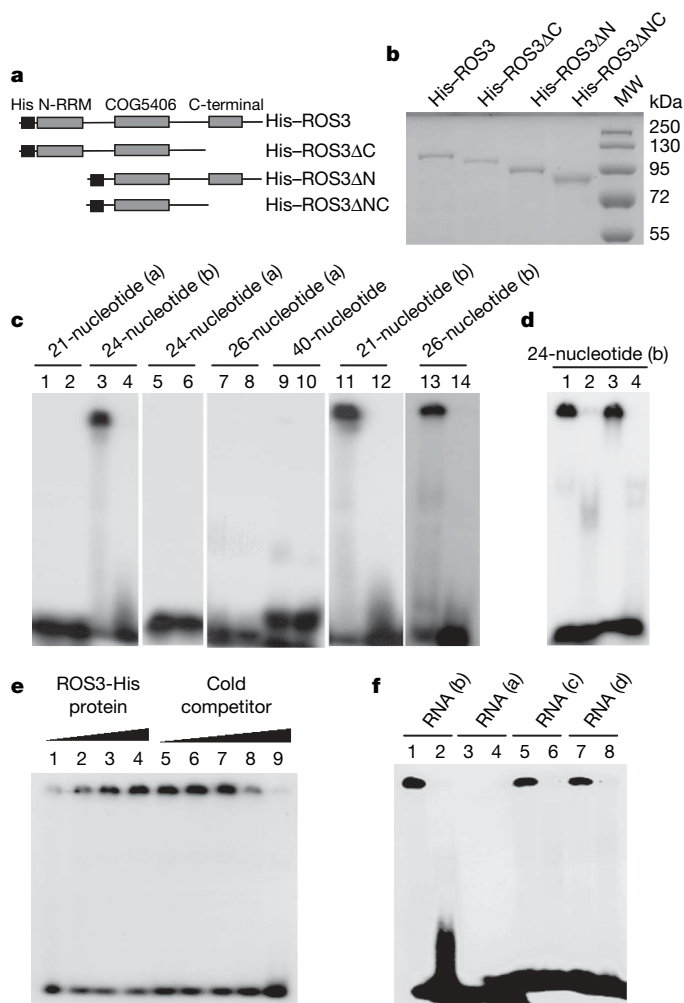


Figure 3 | ROS3 binds small RNAs. **a**, Diagram of ROS3 and its truncated mutant forms. **b**, Coomassie-stained SDS-PAGE gel showing the recombinant proteins used for RNA binding assays. **c**, ROS3 but not ROS3ΔNC binds 21-, 24- and 26-nucleotide single-stranded RNA of sequence b. Lanes 1, 3, 5, 7, 9, 11 and 13, ROS3; lanes 2, 4, 6, 8, 10, 12 and 14, ROS3ΔNC. **d**, ROS3ΔC but not ROS3ΔN or ROS3ΔNC binds 24-nucleotide single-stranded RNA of sequence b. Lane 1, ROS3; lane 2, ROS3ΔN; lane 3, ROS3ΔC; lane 4, ROS3ΔNC. **e**, Protein-concentration-dependent binding to 24-nucleotide single-stranded RNA of sequence b and competition by unlabelled small RNA. Lanes 1–4, increased binding to 24-nucleotide (sequence b) small RNA by increasing ROS3 protein concentration; lanes 5–9, competition by increasing amounts of cold 24-nucleotide RNA (sequence b). **f**, ROS3 binds 25-nucleotide RNA sequences c and d *in vitro*. RNAs of sequences b and a are used as controls. Lanes 1, 3, 5 and 7, ROS3; lanes 2, 4, 6 and 8, ROS3ΔN.

CNG and CNN methylation in *ros3* compared to the wild type in the genomic region corresponding to the small RNAs (Supplementary Fig. 11 and Supplementary Table 2). These results indicate that ROS3 binds to small RNAs *in vivo*, and the small RNAs may direct demethylation of target sequences. Further studies may reveal whether ROS3 is also capable of binding larger RNAs and what specific sequence features it recognizes.

We immunolocalized ROS3 and ROS1 proteins using antibodies recognizing the native proteins or epitope tags fused to ROS1 and ROS3 recombinant proteins. Immunolocalization in *Arabidopsis* leaf nuclei at interphase revealed that in wild-type plants ROS3 is localized in the nucleoplasm as well as the nucleolus (Fig. 4a–c). The ROS3 signals appear as scattered speckle-like structures, ranging in number from 5 to more than 12 per nucleus (Fig. 4 and Supplementary Table 5). This pattern is consistent with the YFP–ROS3 result (Supplementary Fig. 9b, c). ROS1 was similarly found to be dispersed throughout the nucleoplasm and nucleolus, although the immunolocalization signals tended to appear more diffuse and smaller than ROS3 foci (Fig. 4a, b and Supplementary Table 6). In *ros3* and *ros1* mutants, no significant signals could be obtained, which indicates that the antibodies are specific for ROS3 and ROS1, respectively (Fig. 4a). Moreover, epitope-tagged ROS3 and ROS1 proteins displayed interphase localization patterns very similar to the patterns observed for the wild-type proteins (Supplementary Fig. 12). Simultaneous immunolocalization of c-Myc-tagged ROS1 and native ROS3 revealed substantial co-localization of ROS1 and ROS3 (Fig. 4b and Supplementary Table 7).

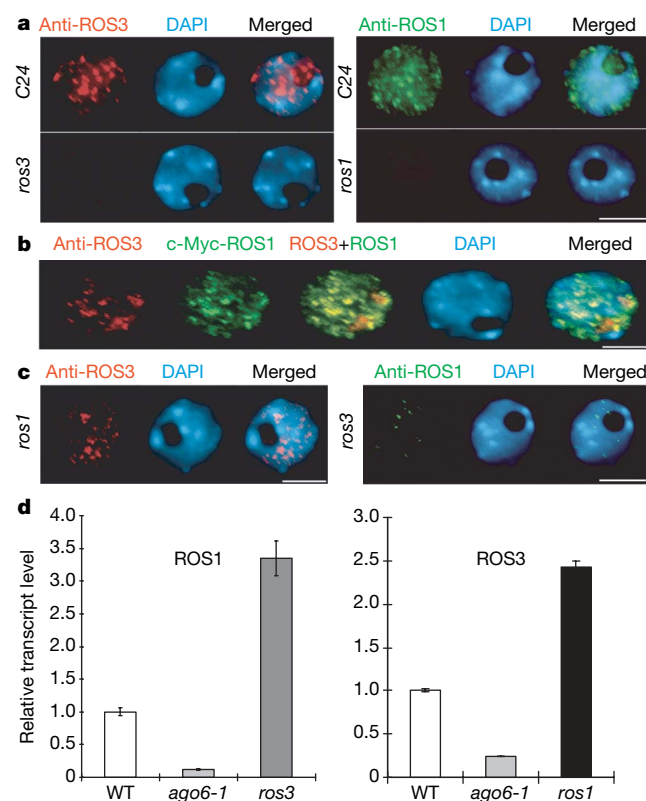


Figure 4 | Co-localization of ROS3 with ROS1 in the nucleus of *Arabidopsis* mesophyll cells and assay of ROS1 and ROS3 mRNA levels. **a**, Localization of ROS3 and ROS1 by their respective antibodies. C24, *Arabidopsis thaliana* C24 ecotype carrying the *RD29A-LUC* transgene (wild type). **b**, Dual immunolocalization of ROS3 and ROS1 with use of anti-ROS3 and anti-Myc. **c**, ROS1 localization in *ros3* mutant and ROS3 localization in *ros1* mutant showing their interdependence for appropriate localization. Scale bars correspond to 5 μ m. **d**, Quantitative RT-PCR assay of relative ROS1 and ROS3 transcript levels in the various genotypes. Error bars represent standard deviation ($n = 3$).

ROS3 and ROS1 immunostaining was also carried out in *ros1* and *ros3* mutants, respectively. ROS1 localization was found to be severely disrupted in the *ros3* mutant background (Fig. 4c). The ROS3 interphase localization pattern was less changed in the *ros1* mutant, although ROS3 nucleolar localization was reduced in the *ros1* mutant (Fig. 4c). These results suggest that ROS1 and ROS3 are interdependent for their nuclear, and especially nucleolar, co-localization.

Because *ROS1* transcript levels are reduced in mutants defective in DNA methylation¹⁶, we tested whether *ros3* may prevent demethylation by reducing *ROS1* expression. *ROS1* messenger RNA level was reduced in *ago6-1* (ref. 17) mutants as expected, but was increased in *ros3* mutants (Fig. 4d). Similarly, *ROS3* mRNA level was decreased in *ago6-1* mutants but increased in *ros1* mutants (Fig. 4d). The results suggest that *ros3* does not reduce *ROS1* mRNA levels and that the DNA methylation status at some critical sites may be sensed to regulate the transcript levels of active DNA demethylation factors such as ROS1 and ROS3.

Our results suggest that ROS3 functions in the same DNA demethylation pathway with ROS1. ROS3 is capable of binding small RNAs *in vitro* and *in vivo*, suggesting that small RNAs and/or larger RNAs bound to ROS3 may guide sequence-specific DNA demethylation by ROS1. In plants, 24-nucleotide siRNAs bound to argonaute 4 (AGO4) and AGO6 can direct DNA methylation and TGS^{5,6,17–21}. Recent studies suggest that piwi-interacting small RNAs function to direct *de novo* DNA methylation and silencing of retrotransposons in mammalian germ cells²². In mammalian cells, promoter-directed siRNAs can also induce TGS^{23,24}. Notably, some promoter-directed small RNAs were recently found to cause gene activation in human cells^{25,26}. Our work showing that a protein required for DNA demethylation is capable of binding to small RNAs contributes to understanding the mechanism of DNA demethylation and adds to the expanding role of small RNAs in gene regulation^{20,21,27}. It also sets the stage for future work to determine the relationship between small RNAs directing DNA methylation and TGS and those directing DNA demethylation and gene activation. Although our data suggest that ROS3 with its bound RNAs may guide sequence-specific demethylation, it is also possible that ROS3 may suppress DNA methylation by sequestering RNAs that guide DNA methylation, or by promoting ROS1 protein stability or nuclear localization.

METHODS SUMMARY

The *ros3* mutant was isolated from a mutagenized population of *RD29A-LUC* plants by luminescence imaging¹⁵. Gene silencing phenotypes of the mutant were examined by RNA blot analysis and nuclear run-on assays. DNA methylation levels were assessed by bisulphite sequencing of genomic DNA or Southern blot analysis. The *ros3* mutant was mapped and cloned following a previously published strategy¹⁰. The subcellular localization of ROS3 protein was determined by confocal imaging of cells or plants expressing YFP-ROS3 (driven by the 35S promoter). The subnuclear co-localization of ROS3 and ROS1 proteins was examined by immunostaining using antibodies against the wild-type protein or epitope tags. Bacterially expressed recombinant ROS3 protein and its truncated forms were purified and used for gel mobility shift assays to examine RNA binding *in vitro*. To test *in vivo* RNA binding by ROS3, ROS3 protein was immunoprecipitated from plant extracts. Small RNAs from the immunoprecipitate were identified by cloning and sequencing.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 11 July; accepted 30 July 2008.

Published online 24 September 2008.

1. Bird, A. DNA methylation patterns and epigenetic memory. *Genes Dev.* **16**, 6–21 (2002).

2. Martienssen, R. A. & Colot, V. DNA methylation and epigenetic inheritance in plants and filamentous fungi. *Science* **293**, 1070–1074 (2001).
3. Tariq, M. & Paszkowski, J. DNA and histone methylation in plants. *Trends Genet.* **20**, 244–251 (2004).
4. Bender, J. DNA methylation and epigenetics. *Annu. Rev. Plant Biol.* **55**, 41–68 (2004).
5. Matzke, M. A. & Birchler, J. A. RNAi-mediated pathways in the nucleus. *Nature Rev. Genet.* **6**, 24–35 (2005).
6. Chan, S. W., Henderson, I. R. & Jacobsen, S. E. Gardening the genome: DNA methylation in *Arabidopsis thaliana*. *Nature Rev. Genet.* **6**, 351–360 (2005).
7. Zhu, J., Kapoor, A., Sridhar, V. V., Agius, F. & Zhu, J. K. The DNA glycosylase/lyase ROS1 functions in pruning DNA methylation patterns in *Arabidopsis*. *Curr. Biol.* **17**, 54–59 (2007).
8. Penterman, J. et al. DNA demethylation in the *Arabidopsis* genome. *Proc. Natl Acad. Sci. USA* **104**, 6752–6757 (2007).
9. Lister, R. et al. Highly integrated single base resolution maps of the epigenome in *Arabidopsis*. *Cell* **133**, 523–536 (2008).
10. Gong, Z. et al. *ROS1*, a repressor of transcriptional gene silencing in *Arabidopsis*, encodes a DNA glycosylase/lyase. *Cell* **111**, 803–814 (2002).
11. Agius, F., Kapoor, A. & Zhu, J. K. Role of the *Arabidopsis* DNA glycosylase/lyase ROS1 in active DNA demethylation. *Proc. Natl Acad. Sci. USA* **103**, 11796–11801 (2006).
12. Morales-Ruiz, T. et al. *DEMETER* and *REPRESSOR OF SILENCING 1* encode 5-methylcytosine DNA glycosylases. *Proc. Natl Acad. Sci. USA* **103**, 6853–6858 (2006).
13. Gehring, M. et al. *DEMETER* DNA glycosylase establishes *MEDEA* polycomb gene self-imprinting by allele-specific demethylation. *Cell* **124**, 495–506 (2006).
14. Kapoor, A., Agius, F. & Zhu, J. K. Preventing transcriptional gene silencing by active DNA demethylation. *FEBS Lett.* **579**, 5889–5898 (2005).
15. Xiong, L. et al. *FIERY1* encoding an inositol polyphosphate 1-phosphatase is a negative regulator of abscisic acid and stress signaling in *Arabidopsis*. *Genes Dev.* **15**, 1971–1984 (2001).
16. Huettel, B. et al. Endogenous targets of RNA-directed DNA methylation and Pol IV in *Arabidopsis*. *EMBO J.* **25**, 2828–2836 (2006).
17. Zheng, X., Zhu, J., Kapoor, A. & Zhu, J. K. Role of *Arabidopsis* AGO6 in siRNA accumulation, DNA methylation and transcriptional gene silencing. *EMBO J.* **26**, 1691–1701 (2007).
18. Wassenaar, M. RNA-directed DNA methylation. *Plant Mol. Biol.* **43**, 203–220 (2000).
19. Vaucheret, H. & Fagard, M. Transcriptional gene silencing in plants: targets, inducers and regulators. *Trends Genet.* **17**, 29–35 (2001).
20. Baulcombe, D. RNA silencing in plants. *Nature* **431**, 356–363 (2004).
21. Brodersen, P. & Voinnet, O. The diversity of RNA silencing pathways in plants. *Trends Genet.* **22**, 268–280 (2006).
22. Aravin, A. A. & Bourc'his, D. Small RNA guides for *de novo* DNA methylation in mammalian germ cells. *Genes Dev.* **22**, 970–975 (2008).
23. Kim, D. H., Villeneuve, L. M., Morris, K. V. & Rossi, J. J. Argonaute-1 directs siRNA-mediated transcriptional gene silencing in human cells. *Nature Struct. Mol. Biol.* **13**, 793–797 (2006).
24. Janowski, B. A. et al. Involvement of AGO1 and AGO2 in mammalian transcriptional silencing. *Nature Struct. Mol. Biol.* **13**, 787–792 (2006).
25. Li, L. C. et al. Small dsRNAs induce transcriptional activation in human cells. *Proc. Natl Acad. Sci. USA* **103**, 17337–17342 (2006).
26. Janowski, B. A. et al. Activating gene expression in mammalian cells with promoter-targeted duplex RNAs. *Nature Chem. Biol.* **3**, 166–173 (2007).
27. Carrington, J. C. & Ambros, V. Role of microRNAs in plant and animal development. *Science* **301**, 336–338 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by National Institutes of Health grants R01GM070795 and R01GM059138 (J.-K.Z.), R01GM077590 and R01GM060380 (C.S.P.), Edward Mallinckrodt Foundation (O.P.) and China Scholarship Council scholarship 2007104542 (F.Z.).

Author Contributions X.Z. did the cloning, mutant analysis, RNA binding and other experiments. J.Z. and A.K. contributed to mutant analysis. D.M., F.Z. and K.I. contributed to DNA methylation analysis. W.-X.L. contributed to small RNA results. O.P. and C.S.P. contributed immunostaining data. J.-K.Z. designed the project and wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to J.-K.Z. (jian-kang.zhu@ucr.edu).

METHODS

Plant materials, mutant isolation and gene cloning. Unless specified otherwise, wild type in this study refers to the *Arabidopsis thaliana* C24 ecotype carrying the *RD29A-LUC* transgene. A T-DNA insertion population was obtained by T-DNA mutagenesis in wild type¹⁵. T2 seedlings were screened for mutants on the basis of luminescence emission. Luminescence imaging was carried out as described²⁸.

RNA analysis. Total RNA was extracted from 15-day-old seedlings by using guanidine thiocyanate (GT) extraction solution. Total RNA and small RNA extraction, and mRNA and small RNA northern blots were as described previously¹⁷. Nuclear run-on assays were carried out as described²⁹. Real-time PCR was carried out as described⁷.

DNA cytosine methylation assay. DNA methylation (by Southern blot or bisulphite sequencing) was analysed on leaves of 4-week-old soil-grown plants and was carried out as described^{11,17,30}. For 5-aza-2'-deoxycytidine (Sigma) treatment, seedlings were grown for 3 weeks on MS agar medium containing 7 µg ml⁻¹ of the drug and then transferred to Whatman paper containing 1/2 MS liquid medium and 300 mM NaCl for 5 h before luciferase imaging.

Positional cloning of *ROS3*. To clone the *ROS3* gene, the homozygous *ros3* mutant in the C24 background with *RD29A-LUC* transgene was crossed to wild-type plants of the Columbia ecotype without the *RD29A-LUC* transgene. The F₂ population was screened for *ros3* mutants based on kanamycin sensitivity of plants containing the *35S-NPTII* transgene and based on leaf morphology. We mapped the *ros3* mutation to the bottom arm of chromosome V on the basis of the *35S-NPTII* transgene silencing phenotype. For fine mapping, we used the leaf developmental phenotype of *ros3*. Using SSLP markers, *ros3* was first mapped to chromosome 5 between markers on *MYN8* and *MTH12*. Through fine mapping the *ros3* mutation was narrowed down to BAC clone K21L19. To identify the *ros3* mutation, candidate genes from wild-type and *ros3* mutant plants were sequenced.

RNA electrophoretic mobility shift assay. *ROS3*-His fusion protein and different truncated *ROS3*-His fusion proteins were cloned into the pET-28a vector and expressed in *Escherichia coli* BL21 cells. The recombinant proteins were purified using Ni-NTA Agarose (Qiagen). Protein-small RNA binding assays were done according to ref. 31 with slight modifications. Briefly, protein binding to RNAs was carried out at room temperature for 30 min in a 30 µl binding mix (10 mM HEPES, pH 7.6, 3 mM MgCl₂, 40 mM KCl, 5% glycerol, 2 mM DTT, 100 nM [γ -³²P]ATP-labelled RNA, and 1 µg *ROS3* protein unless indicated otherwise). To test the effect of varying *ROS3* protein amount on RNA binding, 0.25, 0.5, 0.75 and 1 µg *ROS3* were used in the assays. To test competition by cold RNA, the unlabelled competitor RNA at 100, 300, 500, 2,000 or 10,000 nM was mixed with the labelled RNA before addition of *ROS3* protein. Binding products were resolved on a 3% polyacrylamide native gel.

Immunolocalization. Leaf nuclei used for immunolabelling were isolated from 28-day-old plants as described previously³². Following post-fixation in 4% paraformaldehyde/PBS and blocking with 2% BSA/PBS, the slides were incubated with the following antibodies: rabbit anti-*ROS1*, rabbit anti-*ROS3*, mouse anti-Flag (Sigma) or mouse anti-c-Myc (Sigma). Anti-*ROS1* and anti-*ROS3* antibodies were custom made by Sigma by injecting rabbits with synthetic peptides CLHLPEPFQREQGSE and CRSLDENRQLNGKN, respectively. Secondary antibodies were anti-rabbit or anti-mouse antibodies conjugated to FITC or TRITC (Sigma-Aldrich). In *ros3* mutant plants expressing epitope tags fused to *ROS1* and *ROS3* recombinant proteins, the transgenes (expressed via their respective endogenous promoters) fully rescued the *ros3* and *ros1* mutant

phenotypes (data not shown), which indicates that the recombinant proteins are functional.

Transgenic plants. *ROS3* genomic DNA (including ~1 kb upstream of the initiation codon and 300 bp downstream of the stop codon) was amplified from BAC clone K21L19 with Platinum *pfx* DNA polymerase (Invitrogen) with the forward primer 2711-gDNA-5F and reverse primer 2711-gDNA-3R. The PCR product was cloned between the *SacI* and *HindIII* sites of pCambia 1305.1. *ROS3-flag* and *ROS1-myc* constructs were constructed by inserting *ROS3* genomic DNA with *flag* tag and *ROS1* genomic DNA with *myc* tag into pCambia 1305.1. *ROS3* cDNA was amplified by RT-PCR with forward primer 2711-cDNA-GFP-5F and reverse primer 2711-cDNA-GFP-3R (Supplementary Table 3). The amplified product was cloned into pDONR207 (Invitrogen). pEarlyGate104 was used to express YFP-*ROS3* fusion protein in *Arabidopsis*. For the *ROS3* promoter-GUS construct, the *ROS3* promoter was amplified with forward primer 2711-pro-GUS-5F and reverse primer 2711-pro-GUS-3R. The PCR product was cloned into the *PstI* and *EcoRI* sites of pCambia 1391Z. The constructs were introduced into *Arabidopsis* through floral dip transformation using *Agrobacterium* GV3101. Transgenic lines were selected based on hygromycin resistance for *ros3* complementation, and Basta resistance for YFP-*ROS3* and *ROS3* promoter-GUS expression. GUS staining was as described previously¹⁷.

***ROS3* protein pull-down and small RNA cloning.** The C-terminal portion of *ROS3* protein (amino acids 103–749) was cloned into the pET28a vector (Novagen), expressed in *E. coli*, and purified. Anti-*ROS3* antisera were produced at Calico Biologicals Bio by injecting rabbits with the recombinant protein. Anti-*ROS3* antisera were affinity purified by using the SulfoLink protein kit (Pierce, SulfoLink protein kit number 44995). Two-week-old wild-type *Arabidopsis* seedlings were ground in liquid nitrogen and the extract was used for immunoprecipitation of *ROS3* protein and any RNAs that may be bound to *ROS3* (ref. 33). The immunoprecipitate was fractionated on a denaturing polyacrylamide gel, and putative small RNAs in the 15–30-nucleotide size range were excised from the gel and used for small RNA library construction following established protocols³⁴. PCR products were then cloned into the pGEM-T easy vector system (Promega, number A1360) and sequenced.

28. Ishitani, M., Xiong, L., Stevenson, B. & Zhu, J. K. Genetic analysis of osmotic and cold stress signal transduction in *Arabidopsis*: interactions and convergence of abscisic acid-dependent and abscisic acid-independent pathways. *Plant Cell* **9**, 1935–1949 (1997).
29. Dorweiler, J. E. et al. *mediator of paramutation1* is required for establishment and maintenance of paramutation at multiple maize loci. *Plant Cell* **12**, 2101–2118 (2000).
30. Frommer, M. et al. A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proc. Natl Acad. Sci. USA* **89**, 1827–1831 (1992).
31. Lu, R. et al. Animal virus replication and RNAi-mediated antiviral silencing in *Caenorhabditis elegans*. *Nature* **436**, 1040–1043 (2005).
32. Jasencakova, Z., Meister, A., Walter, J., Turner, B. M. & Schubert, I. Histone H4 acetylation of euchromatin and heterochromatin is cell cycle dependent and correlated with replication rather than with transcription. *Plant Cell* **12**, 2087–2100 (2000).
33. Pontes, O. et al. The *Arabidopsis* chromatin-modifying nuclear siRNA pathway involves a nucleolar RNA processing center. *Cell* **126**, 79–92 (2006).
34. Sunkar, R. & Zhu, J. K. Novel and stress-regulated microRNAs and other small RNAs from *Arabidopsis*. *Plant Cell* **16**, 2001–2019 (2004).

Structural insights into amino acid binding and gene control by a lysine riboswitch

Alexander Serganov^{1*}, Lili Huang^{1*} & Dinshaw J. Patel¹

In bacteria, the intracellular concentration of several amino acids is controlled by riboswitches^{1–4}. One of the important regulatory circuits involves lysine-specific riboswitches, which direct the biosynthesis and transport of lysine and precursors common for lysine and other amino acids^{1–3}. To understand the molecular basis of amino acid recognition by riboswitches, here we present the crystal structure of the 174-nucleotide sensing domain of the *Thermotoga maritima* lysine riboswitch in the lysine-bound (1.9 Å) and free (3.1 Å) states. The riboswitch features an unusual and intricate architecture, involving three-helical and two-helical bundles connected by a compact five-helical junction and stabilized by various long-range tertiary interactions. Lysine interacts with the junctional core of the riboswitch and is specifically recognized through shape-complementarity within the elongated binding pocket and through several direct and K⁺-mediated hydrogen bonds to its charged ends. Our structural and biochemical studies indicate preformation of the riboswitch scaffold and identify conformational changes associated with the formation of a stable lysine-bound state, which prevents alternative folding of the riboswitch and facilitates formation of downstream regulatory elements. We have also determined several structures of the riboswitch bound to different lysine analogues⁵, including antibiotics, in an effort to understand the ligand-binding capabilities of the lysine riboswitch and understand the nature of antibiotic resistance. Our results provide insights into a mechanism of lysine-riboswitch-dependent gene control at the molecular level, thereby contributing to continuing efforts at exploration of the pharmaceutical and biotechnological potential of riboswitches.

RNA sensors play a crucial part in many regulatory loops, owing to their capacity for directing gene expression in response to various stimuli in the absence of protein participation^{6–8}. Recent three-dimensional structures of a thermosensor⁹, a metallosensor¹⁰, a metabolite-bound ribozyme^{11,12} and riboswitches specific for purine nucleobases^{13,14}, and for co-enzymes thiamine pyrophosphate^{15–17} and S-adenosylmethionine^{18,19} have highlighted how each ribosensor uses unique structural features to sense its cognate stimulus. However, the molecular details of the organization of amino-acid-specific riboswitches, such as the lysine riboswitch, which efficiently discriminates against other free amino acids, their precursors and amino acids within a peptide context^{1,2,5}, remain obscure. The determination of the lysine riboswitch structure presents a considerable challenge, because of its large metabolite-sensing domain, predicted to form a five-way junction^{1–3}.

The *T. maritima* riboswitch is a typical lysine riboswitch^{1–3,5} (Supplementary Figs 1 and 2a) which uses a transcriptional attenuation mechanism to repress the production of aspartate-semialdehyde dehydrogenase³, which is involved in the synthesis of a precursor for

methionine, threonine, lysine and diaminopimelate. The structure of the lysine-bound riboswitch domain, also known as an 'L box', features three-helical and two-helical bundles radiating from a compact five-helical junction (Fig. 1a–c and Supplementary Fig. 2b–d) that contains lysine inserted into its core. The junction is organized on the basis of a modified four-way junction²⁰ through colinear stacking of helices P1 and P2, and helices P4 and P5, positioned as two intersecting lines of an uneven letter 'X'. The P2–P2a–L2 stem-loop reverses its orientation through two turns important for riboswitch function^{5,21}: one of them adjacent to a loop E motif^{22,23} and the other centred on a turn that replaces the kink-turn motif²⁴ found in other lysine riboswitches (Supplementary Fig. 3). Stems P2 and P3 are aligned by an unusual kissing-loop complex between loops L2 and L3 (Fig. 1d and Supplementary Fig. 4), whereas parallel stems P2 and P4 are anchored by a conserved loop (L4)–helix (P2) interaction (Fig. 1e).

The five-helical junction contains three layers of nucleotides, each composed of two interacting base pairs, organized around the centrally positioned lysine which fits into a tight pocket and is specifically recognized by its charged ends (Fig. 2a–c). This lysine-bound pocket architecture (Fig. 2) is also retained in lysine analogue complexes (Fig. 3) outlined later. The top layer, composed of G14–C78 and G115–C139 base pairs (Fig. 2d), is stabilized through ribose zipper and type I A-minor triple interactions with invariant A81, which in turn is paired with Na⁺-bound G80 (Fig. 2d and Supplementary Fig. 5a). The middle layer, containing G12–C79 and G114–U140 base pairs, forms specific interactions with the bound lysine (Fig. 2e). Both the carboxylate and ammonium groups of the lysine 'main chain' segment are hydrogen-bonded to the minor groove edges of purine bases and sugar 2'-OH groups (Fig. 2b, e), whereas the ε-ammonium protons of the side chain form hydrogen bonds with a non-bridging phosphate oxygen, a sugar ring oxygen and a tightly bound water molecule, W1 (Fig. 2c). In addition, lysine is sandwiched between the A81 base (Fig. 2d) and the G11•G163 base pair from the bottom junctional layer (Fig. 2e and Supplementary Fig. 5b), thereby stabilizing the top of the P1 helix which is necessary for gene expression control. The bound lysine facilitates the holding together of the stacked P1–P2 and P4–P5 helical segments, and contributes to the positioning of the P3 helix by locking up G80, which is placed over the ε-ammonium group of lysine, and stacks with the A82–U113 base pair of P3 (Fig. 2a). This stable junctional conformation is further reinforced by a tertiary stacking interaction between G110 and A164 (Fig. 2a) characteristic for riboswitches from thermophiles, as well as other interhelical contacts (Figs 1b, c and 2a).

A notable feature of the lysine-binding pocket is a K⁺ cation (Fig. 2b), which binds a carboxyl oxygen of lysine and zippers up the junction using several coordination bonds. The K⁺ cation is directly observed on the anomalous map (Supplementary Fig. 6)

¹Structural Biology Program, Memorial Sloan-Kettering Cancer Center, New York, New York 10065, USA.

*These authors contributed equally to this work.

and is replaceable by its mimics, Cs^+ and Tl^+ , but not by Mn^{2+} (Supplementary Figs 7–9), a mimic of Mg^{2+} (refs 25 and 26). The importance of K^+ for lysine binding has been demonstrated in primer extension experiments. In the presence of lysine and at physiological concentration of K^+ , reverse transcriptase pauses before the junction at A169, reflecting the formation of a stable junctional

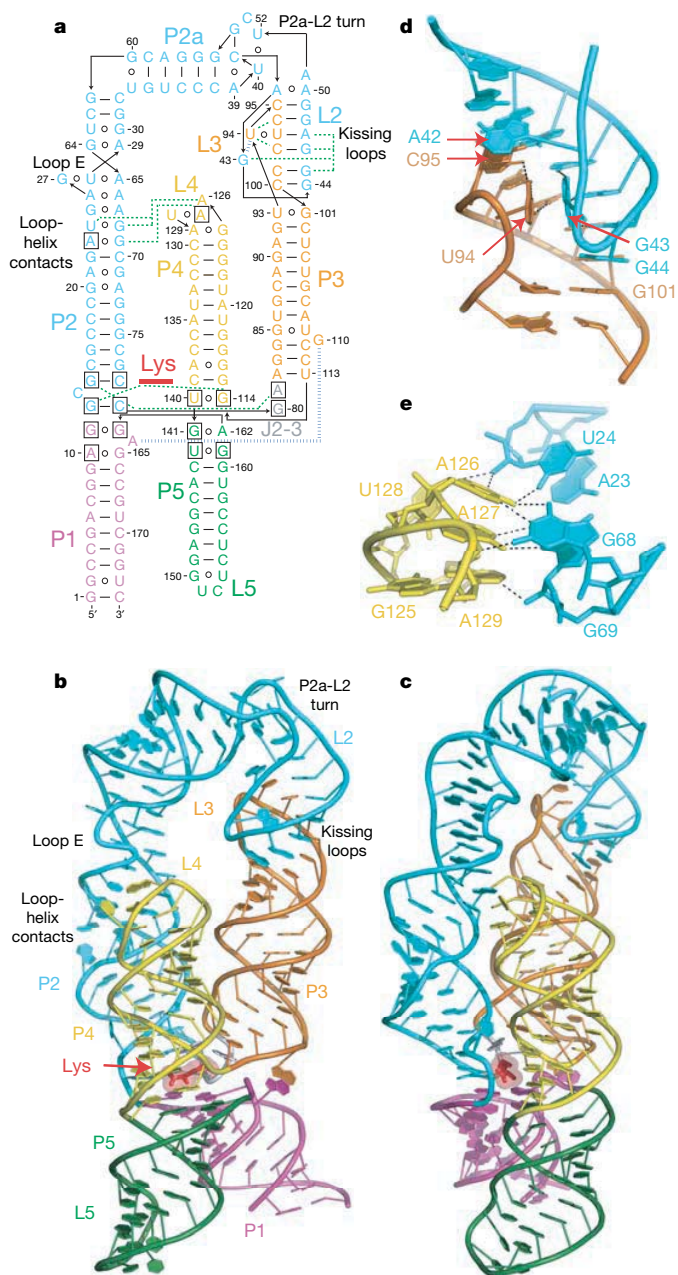


Figure 1 | Overall structure and long-range tertiary interactions of the lysine-bound *T. maritima* riboswitch. **a**, Schematic of the riboswitch fold observed in the crystal structure of the complex. The bound lysine is in red. The RNA domains are depicted in colours used for subsequent figures. Base-specific tertiary contacts and long-range stacking interactions are shown as thin green and thick blue dashed lines, respectively. Nucleotides invariant in known lysine riboswitches are boxed. **b**, **c**, Overall lysine riboswitch structure in a ribbon representation showing front (**b**) and rotated by ~60° (**c**) views. **d**, The L2–L3 kissing loop interaction is formed by six base pairs, supplemented by interstrand stacking interactions between A42 and C95, G43 and U94, and G44 and G101. Hydrogen bonds between interstrand base pairs and orthogonally aligned G43 and U94 bases are depicted by dashed lines. **e**, The L4-loop–P2-helix interaction formed by an insertion of the A126–A127–A129 stack of L4 into the RNA groove of P2 distorted by non-canonical base pairs.

conformation (Fig. 4a, lanes 8, 10). The pause decreases 33-fold after the replacement of K^+ by Na^+ (Fig. 4a, lanes 2, 4), suggesting that there is either reduced stability or failure to generate a junction competent for lysine recognition under K^+ -free conditions. These results have been supported by equilibrium dialysis experiments with *T. maritima* and *Bacillus subtilis* riboswitches. In both cases, lysine binding affinity significantly decreased when K^+ was omitted or replaced by Na^+ (Fig. 4b). Note that lysine binds better to a riboswitch from a thermophile than from a mesophile.

The preference of K^+ for binding to the negatively charged carboxylate group contrasts with Mg^{2+} -mediated phosphate recognition in other ribosensors^{11,12,15,16} and might also be a characteristic for other amino-acid-specific riboswitches. Because more than 20 lysine-binding proteins (listed in the Protein Data Bank) do not use cations to mediate lysine recognition, this feature is probably unique to RNA, given that it lacks the positively-charged side chains found in proteins.

Folding of most RNAs, including the *B. subtilis* lysine riboswitch²¹, requires Mg^{2+} . However, the crystals of the *T. maritima* lysine riboswitch can be grown in the absence of Mg^{2+} (Supplementary Fig. 10). Moreover, a Mn^{2+} soak does not replace cations in the structure grown with Mg^{2+} , and, on the basis of the coordination distances, most of these cations can be assigned as Na^+ . These results indicate that the L box from a thermophile does not critically depend on Mg^{2+} , the function of which can be co-opted by monovalent cations.

The identification of antibiotic-resistance mutations^{27,28} in the lysine riboswitch^{1,2}, together with the demonstration of a direct interaction between the riboswitch and lysine-like antibacterial compounds^{1,2,5}, suggest that riboswitch targeting, along with other processes²⁹, is an important component of the antibiotic activity. To understand the molecular basis of antibiotic resistance and explore the pharmaceutical potential of the lysine riboswitch, we have determined structures of the riboswitch bound to antibacterial compounds S-(2-aminoethyl)-L-cysteine (AEC) and L-4-oxalysine⁵, which contain sulphur and oxygen at position C4, respectively (Fig. 3a). Because the pocket has a small cavity between the C4 and N7 positions of bound lysine (Fig. 3c and Supplementary Fig. 11a), both C4-substituted analogues can be placed within the pocket in a manner similar to bound lysine (Fig. 3b, top panel), suggesting the potential for incorporation of even larger C4-substituents. Despite similar placement within the pocket, primer extension assays suggest that there is weaker binding of AEC and L-4-oxalysine to the riboswitch (Fig. 4c and Supplementary Fig. 12), possibly due to an increased electronegativity of substituents at the C4 position.

Next, we determined the structures with lysine analogues L-homoarginine and N6-1-iminoethyl-L-lysine, where the ϵ -ammonium group of lysine is replaced by a guanidinium group and its methyl-substituted variant, respectively (Fig. 3a)⁵. In these structures (Fig. 3b), the side chains of lysine analogues are slightly shifted to provide better stacking interactions with the G80 base. The nitrogen atoms of the guanidinium-like extensions replace water molecules W1 and W2, found in the lysine complex (Fig. 2b). The G163 sugar is slightly rotated, so that the hydrogen bond pattern of W1 is retained by both ligands, whereas L-homoarginine forms extra hydrogen bonds with G163. Although most RNA-ligand contacts are preserved, both analogues demonstrate weaker interactions than lysine in primer extension (Fig. 4c and Supplementary Fig. 12) and in-line probing⁵ experiments, emphasizing the fine structural complementarity between the ϵ -ammonium group of lysine and RNA.

The lysine-binding pocket has two openings, which could be exploited for the design of next generation lysine-like analogues (Fig. 3c, d and Supplementary Fig. 11b, c). One of the openings could accommodate modifications or extensions of the carboxylate group, possibly by substituting the K^+ cation. The other smaller opening could allow extensions from N9 of L-homoarginine and iminoethyl-L-lysine. The analogue-bound structures and the biochemical data⁵ indicate that the lysine-binding pocket is rather rigid, and only

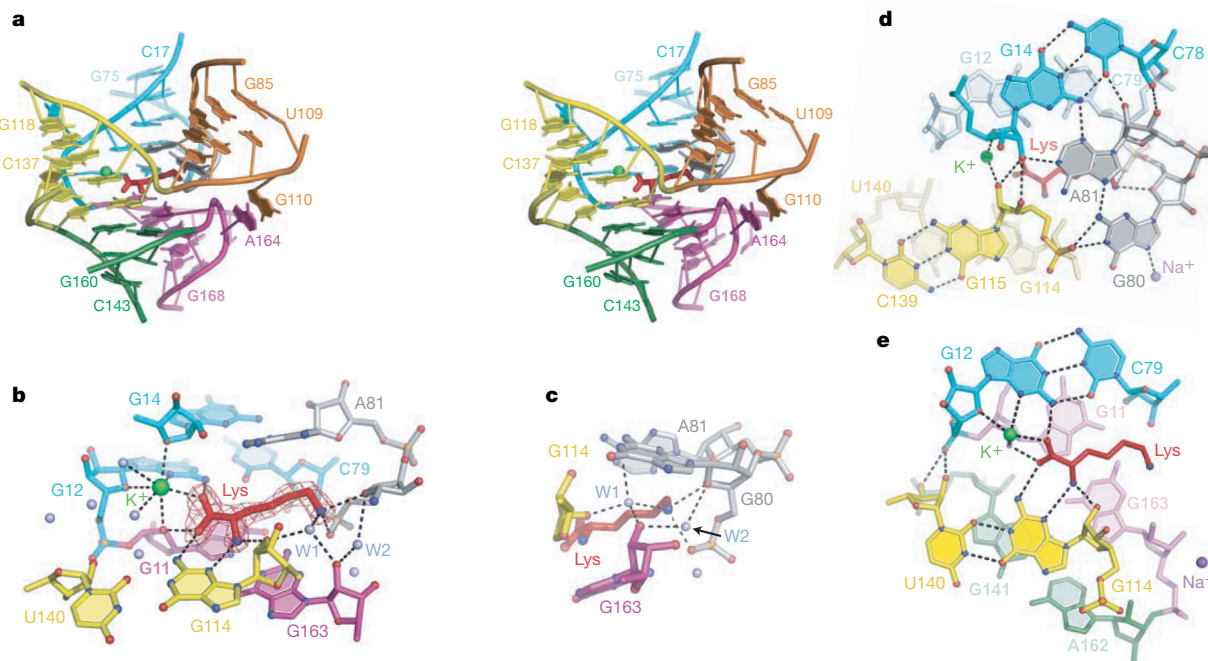


Figure 2 | Structure and interactions in the junctional region of the lysine riboswitch. **a**, Stereo view of the junction with bound lysine. Green sphere depicts a K^+ cation. **b**, Details of riboswitch lysine interactions. Lysine is positioned within the omit $F_o - F_c$ electron density map contoured at 3.5σ

accommodates compounds which can sterically fit the pocket. Therefore, lysines in a polypeptide chain and branched amino acids are not recognized by the riboswitch. The lysine riboswitch also

level. Water molecules are shown as light blue spheres. K^+ cation coordination and hydrogen bonds are depicted by dashed lines. **c**, Direct and water-mediated interactions involving ϵ -ammonium group of lysine. **d**, **e**, Interactions in the top (**d**) and middle (**e**) junctional layers.

discriminates against smaller amino acids that fit into the pocket (data not shown) but are unable to make essential intermolecular contacts in the vicinity of G80.

To gain insights into lysine-induced conformational rearrangements of the riboswitch in solution, we performed footprinting experiments using in-line probing, specific for flexible RNA regions, and cleavage by the nucleases T2 (single-stranded RNA) and V1 (paired and stacked regions). As in *B. subtilis*^{1,5}, the transition from the free to the lysine-bound states of the *T. maritima* riboswitch is accompanied by conformational changes within the junctional core, detected as strong cleavage reductions in both in-line and V1 probing (Fig. 4d–f and Supplementary Fig. 13). However, the kissing loop complex is probably preformed in the free riboswitch^{5,21}, as evident from the absence of cleavages in in-line^{1,5} (Fig. 4d) and T2 probing (Supplementary Fig. 13b), coupled with V1 cleavages at nucleotides 45–46 (Fig. 4e), which are only weakly enhanced by lysine binding. The P2–L4 tertiary contact seems to be more dynamic in character as reflected in the rather strong T2 (nucleotides 126–127, Fig. 4f) cleavage patterns in the lysine-free form, coupled with only weak protection on complex formation. The projection of the in-line cleavage reductions (this study and ref. 1) on the structure (Fig. 4g) provides the first glimpses into the primary determinants of lysine recognition and how the riboswitch junction folds on ligand binding. Assuming that helix P1 is not fully formed before lysine binding and that protections in P5 are, at least in part, due to P1–P5 interactions, we propose that the main structural changes in solution, seen in both studies, involve stabilization of G80 and formation of the G12–C79

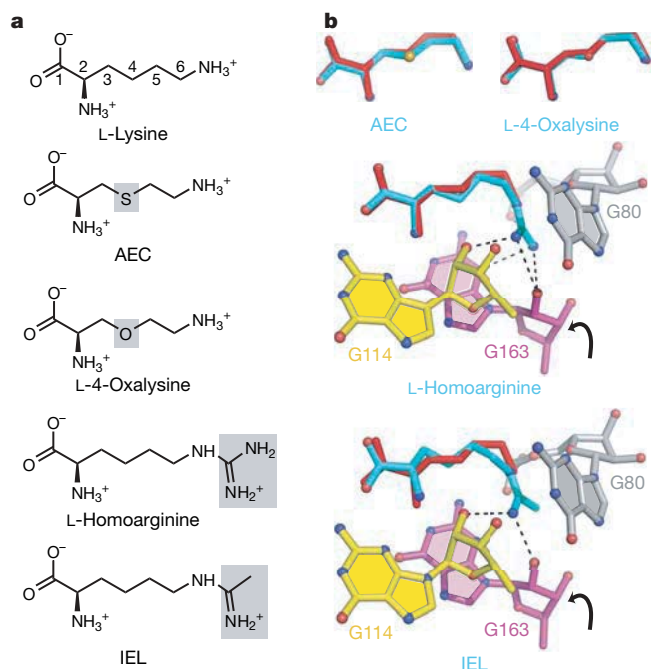


Figure 3 | Interactions of lysine analogues with the riboswitch. **a**, Chemical structures of lysine and its analogues. **b**, Conformation and interactions of lysine analogues with the riboswitch. Lysine (red) and analogues (cyan) are superposed. Interactions between riboswitch and lysine analogues should be compared with lysine recognition in Fig. 2b. **c**, Cross-section through the surface view of the lysine-binding pocket showing the opening next to the carboxyl group of lysine (red arrow) and the free space next to the C4 atom (blue arrow). **d**, Cross-section through the homoarginine-riboswitch complex showing the opening (red arrow) next to the guanidinium group.

and G11•G163 junctional base pairs, followed by stabilization of the surrounding regions and the P1 helix (Supplementary Fig. 14).

Prompted by pre-formation of tertiary riboswitch elements in solution, we have crystallized the riboswitch in the absence of lysine. This 3.1 Å structure (Supplementary Fig. 15) is very similar to the lysine-bound form, except that it lacks lysine and junctional K⁺.

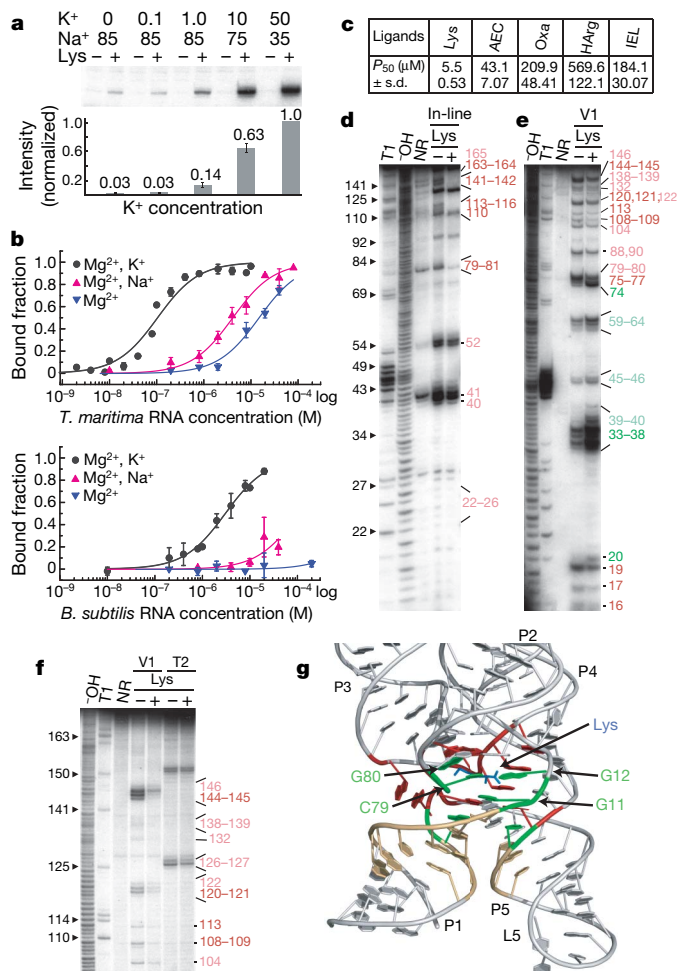


Figure 4 | Probing lysine riboswitch tertiary structure. **a**, Primer extension analysis at various K⁺ concentrations. ³²P-labelled oligonucleotide was annealed to the 3' end of 265-nucleotide *T. maritima* RNA (1 μM) and extended by reverse transcriptase in the presence of 2 mM MgCl₂, at indicated concentrations of monovalent cations (in mM), and with or without a tenfold excess of lysine over RNA. **b**, Lysine binding affinity measured by equilibrium dialysis for riboswitches from *T. maritima* (top) and *B. subtilis* (bottom). Mg²⁺, K⁺ and Na⁺ concentrations are 20, 100 and 100 mM, respectively. The dissociation constants (mean ± s.d., μM; *n* = 2–4) are: *T. maritima*, 0.10 ± 0.03 (Mg²⁺ + K⁺), 4.14 ± 0.67 (Mg²⁺ + Na⁺), 15.93 ± 0.09 (Mg²⁺); *B. subtilis*, 2.95 ± 0.30 (Mg²⁺ + K⁺). **c**, The apparent ligand concentration at which reverse transcriptase pausing is half-maximally attained (*P*₅₀; *n* = 2) in the primer extension experiments. HArg, L-homoarginine; IEL, iminoethyl-L-lysine; Lys, L-lysine; Oxa, L-4-oxalysine. **d**, In-line probing of 5' ³²P-labelled *T. maritima* 174-nucleotide RNA (1.6 nM) in the absence and presence of lysine (1.1 μM). T1 and -OH designate RNase T1 and alkaline ladders, respectively. NR, no reaction. Strong and weak cleavage reductions are shown in red and pink colours, respectively. **e**, **f**, Probing of 174-nucleotide RNA (1 μM) by V1 and T2 nucleases with or without a tenfold excess of lysine. Weak and strong cleavage enhancements are shown in light and dark green, respectively, in **e**. **g**, Strong lysine-induced cleavage reductions are colour coded in the riboswitch structure. Light orange, green and red are reductions identified in ref. 1 using *B. subtilis* RNA with a short P1 helix, overlapping reductions of the present study and in ref. 1, and extra reductions found in the present study, respectively.

Although the absence of the expression platform and long P1 helix facilitate formation of this conformation, the structure emphasizes the importance of RNA interactions in maintaining the riboswitch conformation, suggests a crucial role of K⁺ in mediation of lysine-RNA but not RNA–RNA interactions, and reinforces the feasibility of lysine stabilizing a largely preformed riboswitch structure.

The L box structure readily explains mutations that deregulate gene expression and confer resistance to AEC^{27,28} (Supplementary Fig. 16). The G12A, G12C and G81A mutations disrupt the lysine-binding pocket, whereas the G11A, G11U, G9C and C166U substitutions prevent pairing of the P1 helix. Therefore, intracellular lysine and AEC cannot bind the mutated riboswitches, and the segment downstream of G161 engages in formation of an anti-terminator stem (Supplementary Fig. 1), resulting in constitutive lysine production.

The unusual architecture and high ligand specificity, achieved through a combination of shape complementarity and K⁺-assisted recognition of the bound lysine, distinguishes the lysine riboswitch from other riboswitches. Given the importance of lysine riboswitch-controlled gene expression for bacterial viability and the absence of the diaminopimelate pathway in mammals, the structure provides critical details towards facilitating the design of lysine-like analogues targeting riboswitches and other cellular sites.

METHODS SUMMARY

Crystallization and structure determination. A 0.4 mM lysine riboswitch complex was prepared by mixing *in vitro* transcribed RNA and lysine in a buffer containing 100 mM potassium acetate, pH 6.8, and 4 mM MgCl₂. Crystals were grown by hanging-drop vapour diffusion after mixing the complex and the reservoir (18% (w/v) PEG4000, 100 mM sodium citrate, pH 5.7, and 20% isopropanol (v/v)) solutions at 1:1 ratio. For soaking, crystals were placed in the reservoir solution with PEG4000 replaced by 20% PEG400, and then incubated in the presence of either 3 mM [Ir(NH₃)₆]³⁺ or 10–50 mM of Cs⁺, Tl⁺ and Mn²⁺ salts for 7 h. Crystals were flash frozen in liquid nitrogen and data were collected at 100 K. The structure was determined using 2.4 Å multiwavelength anomalous dispersion iridium data and SHARP³⁰. The RNA model was built using TURBO-FRODO (<http://www.afmb.univ-mrs.fr/~TURBO->), and refined using the 1.9 Å native data set to *R*_{work}/*R*_{free} 19.2/22.9 (Supplementary Table 1 and Supplementary Fig. 17). Lysine and cations were added to the model on the basis of analysis of 2*F*_o − *F*_c, *F*_o − *F*_c and anomalous electron density maps. Cations were modelled on the basis of the number of coordination bonds, their distances, coordination geometry and temperature factors. Analogue-bound and metal-soaked riboswitch structures were refined using the native riboswitch model (Supplementary Tables 2 and 3). Figures were prepared with PyMol (<http://www.pymol.org>).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 1 May; accepted 6 August 2008.

Published online 10 September 2008.

- Sudarsan, N., Wickiser, J. K., Nakamura, S., Ebert, M. S. & Breaker, R. R. An mRNA structure in bacteria that controls gene expression by binding lysine. *Genes Dev.* 17, 2688–2697 (2003).
- Grundy, F. J., Lehman, S. C. & Henkin, T. M. The L box regulon: lysine sensing by leader RNAs of bacterial lysine biosynthesis genes. *Proc. Natl Acad. Sci. USA* 100, 12057–12062 (2003).
- Rodionov, D. A., Vitreschak, A. G., Mironov, A. A. & Gelfand, M. S. Regulation of lysine biosynthesis and transport genes in bacteria: yet another RNA riboswitch? *Nucleic Acids Res.* 31, 6748–6757 (2003).
- Mandal, M. *et al.* A glycine-dependent riboswitch that uses cooperative binding to control gene expression. *Science* 306, 275–279 (2004).
- Blount, K. F., Wang, J. X., Lim, J., Sudarsan, N. & Breaker, R. R. Antibacterial lysine analogs that target lysine riboswitches. *Nature Chem. Biol.* 3, 44–49 (2007).
- Serganov, A. & Patel, D. J. Ribozymes, riboswitches and beyond: regulation of gene expression without proteins. *Nature Rev. Genet.* 8, 776–790 (2007).
- Nudler, E. & Mironov, A. S. The riboswitch control of bacterial metabolism. *Trends Biochem. Sci.* 29, 11–17 (2004).
- Winkler, W. C. & Breaker, R. R. Regulation of bacterial gene expression by riboswitches. *Annu. Rev. Microbiol.* 59, 487–517 (2005).
- Chowdhury, S., Maris, C., Allain, F. H. & Narberhaus, F. Molecular basis for temperature sensing by an RNA thermometer. *EMBO J.* 25, 2487–2497 (2006).
- Dann, C. E. III *et al.* Structure and mechanism of a metal-sensing regulatory RNA. *Cell* 130, 878–892 (2007).

11. Klein, D. J. & Ferre-D'Amare, A. R. Structural basis of *glmS* ribozyme activation by glucosamine-6-phosphate. *Science* **313**, 1752–1756 (2006).
12. Cochrane, J. C., Lipchick, S. V. & Strobel, S. A. Structural investigation of the *GlmS* ribozyme bound to its catalytic cofactor. *Chem. Biol.* **14**, 97–105 (2007).
13. Batey, R. T., Gilbert, S. D. & Montange, R. K. Structure of a natural guanine-responsive riboswitch complexed with the metabolite hypoxanthine. *Nature* **432**, 411–415 (2004).
14. Serganov, A. *et al.* Structural basis for discriminative regulation of gene expression by adenine- and guanine-sensing mRNAs. *Chem. Biol.* **11**, 1729–1741 (2004).
15. Serganov, A., Polonskaia, A., Phan, A. T., Breaker, R. R. & Patel, D. J. Structural basis for gene regulation by a thiamine pyrophosphate-sensing riboswitch. *Nature* **441**, 1167–1171 (2006).
16. Thore, S., Leibundgut, M. & Ban, N. Structure of the eukaryotic thiamine pyrophosphate riboswitch with its regulatory ligand. *Science* **312**, 1208–1211 (2006).
17. Edwards, T. E. & Ferre-D'Amare, A. R. Crystal structures of the thi-box riboswitch bound to thiamine pyrophosphate analogs reveal adaptive RNA-small molecule recognition. *Structure* **14**, 1459–1468 (2006).
18. Montange, R. K. & Batey, R. T. Structure of the *S*-adenosylmethionine riboswitch regulatory mRNA element. *Nature* **441**, 1172–1175 (2006).
19. Gilbert, S. D., Rambo, R. P., Van Tyne, D. & Batey, R. T. Structure of the SAM-II riboswitch bound to *S*-adenosylmethionine. *Nature Struct. Mol. Biol.* **15**, 177–182 (2008).
20. Weixlbaumer, A. *et al.* Crystal structure of the ribosome recycling factor bound to the ribosome. *Nature Struct. Mol. Biol.* **14**, 733–737 (2007).
21. Blouin, S. & Lafontaine, D. A. A loop-loop interaction and a K-turn motif located in the lysine aptamer domain are important for the riboswitch gene regulation control. *RNA* **13**, 1256–1267 (2007).
22. Correll, C. C., Freeborn, B., Moore, P. B. & Steitz, T. A. Metals, motifs, and recognition in the crystal structure of a 5S rRNA domain. *Cell* **91**, 705–712 (1997).
23. Leontis, N. B. & Westhof, E. The 5S rRNA loop E: chemical probing and phylogenetic data versus crystal structure. *RNA* **4**, 1134–1153 (1998).
24. Klein, D. J., Schmeing, T. M., Moore, P. B. & Steitz, T. A. The kink-turn: a new RNA secondary structure motif. *EMBO J.* **20**, 4214–4221 (2001).
25. Basu, S. *et al.* A specific monovalent metal ion integral to the AA platform of the RNA tetraloop receptor. *Nature Struct. Biol.* **5**, 986–992 (1998).
26. Feig, A. L. & Uhlenbeck, O. C. in *The RNA World Second Edition* (eds Gesteland, R. F., Cech, T. R. & Atkins, J. F.) 287–319 (Cold Spring Harbor Laboratory Press, 1999).
27. Lu, Y., Shevtchenko, T. N. & Paulus, H. Fine-structure mapping of *cis*-acting control sites in the *lysC* operon of *Bacillus subtilis*. *FEMS Microbiol. Lett.* **92**, 23–27 (1992).
28. Patte, J. C., Akrim, M. & Mejean, V. The leader sequence of the *Escherichia coli lysC* gene is involved in the regulation of *LysC* synthesis. *FEMS Microbiol. Lett.* **169**, 165–170 (1998).
29. Ataide, S. F. *et al.* Mechanisms of resistance to an amino acid antibiotic that targets translation. *ACS Chem. Biol.* **2**, 819–827 (2007).
30. de La Fortelle, E. & Bricogne, G. in *Methods in Enzymology* 472–494 (Academic Press, 1997).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank personnel of beamline X29 at the Brookhaven National Laboratory and beamlines 24-ID-C/E at the Advanced Photon Source, Argonne National Laboratory, funded by the US Department of Energy. We thank O. Ouerfelli for the synthesis of iridium hexamine. D.J.P. was supported by funds from the National Institutes of Health.

Author Contributions L.H. crystallized the *T. maritima* lysine riboswitch; A.S. determined the structures and was assisted by L.H. during refinement; A.S. and L.H. performed biochemical experiments; and A.S. and D.J.P. wrote the manuscript. All authors discussed the results and commented on the manuscript.

Author Information Atomic coordinates of the X-ray structures of the lysine riboswitch bound to ligands have been deposited in the RCSB Protein Data Bank under the following accession codes: lysine, 3DIL; AEC, 3DIG; L-4-oxalysine, 3DJ0; homoarginine, 3DIQ; and iminoethyl-L-lysine, 3DIR. Codes for other structures are: free state, 3DIS; [Ir(NH₃)₆]³⁺-soaked, 3DIO; Cs⁺-soaked, 3DIM; Tl⁺-soaked, 3DJ2; Mn²⁺-soaked, 3DIY; K⁺-anomalous, 3DIX; and Mg²⁺-free form, 3DIZ. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to A.S. (serganoa@mskcc.org) or D.J.P. (pateld@mskcc.org).

METHODS

RNA preparation and complex formation. The lysine riboswitch, followed by the hammerhead ribozyme, was transcribed *in vitro* using T7 RNA polymerase. RNA was purified by denaturing polyacrylamide gel electrophoresis (PAGE) and anion-exchange chromatography. Lysine analogues were added to RNA at a 2.5–2.75 to 1 molar ratio. To form a complex without Mg^{2+} , the RNA was mixed with 100 mM potassium-acetate, 1.0 mM EDTA and lysine. To prepare RNA for crystallization without lysine, 0.2 mM RNA was supplemented with 50 mM potassium acetate, pH 6.8, 50 mM sodium acetate, pH 6.9, and 2 mM MgCl_2 , and concentrated twofold by Speedvac. Before crystallization, sodium-citrate, pH 5.7, was added to the mixture up to 100 mM, and the RNA sample was heated at 55 °C for 5 min and cooled on ice for 15 min.

Crystallization. Hanging drops were prepared by mixing 1 μl of the complex with 1 μl of the reservoir solution. The drops were equilibrated against 1 ml of reservoir solution at 20 °C for ~1–2 weeks. The riboswitch in the free state was crystallized in the solution containing 21% (w/v) PEG4000, 100 mM Bis-Tris, pH 5.5, and 25% isopropanol (v/v). For cryoprotection, crystals were quickly passed through the stabilizing solution, which was the reservoir solution with PEG4000 replaced by 20% PEG400. For soaking, crystals were passed through several 5 μl drops of the stabilizing solution, and then incubated in 5 μl of stabilizing solution supplemented with 3 mM $[\text{Ir}(\text{NH}_3)_6]\text{Cl}_3$, 10 mM CsCl, 10 mM thallium acetate, or 50 mM MnCl_2 salts for 7–8 h.

X-ray crystallography. Data were reduced using HKL2000 (HKL Research). The structure was determined using the autoSHARP option of SHARP and 2.4 Å MAD iridium data. The resulting experimental map was of excellent quality (Supplementary Fig. 17a) for most of the RNA molecule. The structure contains 1 RNA and 16 iridium hexamine sites (2 of them are split) per asymmetric unit (Supplementary Fig. 17b). The RNA model was built using 2.4 Å MAD electron density map, and then refined with REFMAC³¹ using 1.9 Å native data. The bound lysine and several cations with octahedral coordination geometry were added on the basis of the $2F_o - F_c$ and $F_o - F_c$ electron density maps. Cations were interpreted as Na^+ or Mg^{2+} on the basis of the coordination distances in the range of 2.25–2.85 Å (Supplementary Fig. 17d) and 2.0–2.3 Å, respectively. Water molecules were added using ARP/wARP³². K^+ cations were added on the basis of the 1.9 Å $2F_o - F_c$ and 2.9 Å anomalous (Supplementary Fig. 6) electron density maps and typical K^+ coordination distances (Supplementary Fig. 17c). The final 1.9 Å riboswitch model contains 174 nucleotides, 1 lysine molecule, 1 magnesium cation, 3 potassium and 29 sodium cations. Cs^+ , Ti^+ , and Mn^{2+} cations were modelled on the basis of the anomalous maps (Supplementary Figs 7–9), whereas addition of other cations was guided by the high-resolution structure and the analysis of coordination geometries and distances.

Primer extension assay. Primer extension experiments were performed using 265-nucleotide full-length riboswitch. The ^{32}P -labelled 13-mer DNA oligonucleotide (100,000 c.p.m.), complementary to nucleotides 253–265 of RNA, was annealed to the 3' end of RNA (final RNA concentration 1 μM in the assay).

Primer extension was conducted in 15 μl volume with 40 U of moloney murine leukaemia virus reverse transcriptase in 50 mM Na-HEPES, pH 7.9, 2 mM MgCl_2 , and variable concentrations of NaCl and KCl (Fig. 4), with or without a tenfold excess of lysine over RNA. After 30 min incubation at 37 °C, the reactions were precipitated with ethanol, dissolved in loading buffer and analysed by 10% PAGE. Efficiency of lysine-induced pausing of reverse transcriptase at nucleotide A169 was quantified using FLA-7000 PhosphorImager and Image Gauge software (Fujifilm). Band intensities from independent experiments were averaged after gel-loading correction, background subtraction and normalization.

The primer extension assay in the presence of lysine analogues was performed in 50 mM Na-HEPES, pH 7.9, 2 mM MgCl_2 , 50 mM KCl, and lysine analogues in the range 10^{-8} – 10^{-2} M. The data were fitted using a bimolecular equilibrium equation (Supplementary Fig. 12) and the resulting P_{50} values are reported in Fig. 4c. Reverse transcriptase sequencing reactions were run in parallel.

Footprinting studies. For footprinting experiments, the 174-nucleotide metabolite-sensing domain of the riboswitch was radioactively labelled at the 5' end by the kinase reaction. For in-line probing, 30,000–300,000 c.p.m. of 174-nucleotide RNA (1.6–16 nM) was incubated in 10–30 μl solution containing 50 mM Tris-HCl, pH 8.3, 100 mM KCl and 20 mM MgCl_2 in the absence or presence of ~6–600-fold excess of lysine (Fig. 4) at room temperature for ~40 h. After incubation, aliquots were analysed by PAGE along with alkaline ladder and T1 nuclease digestion.

For nuclease footprinting experiments, 20 μl samples of 174-nucleotide RNA (100,000 c.p.m.) with a final RNA concentration 1 μM were preheated at 37 °C for 10 min in 50 mM Na-HEPES, pH 7.9, 50 mM KCl and 2 mM MgCl_2 . Mixtures were incubated with a tenfold excess of lysine over RNA at 37 °C for 15 min. Cleavage reactions were performed with 0.0025 U RNase V1 (Pierce) or 0.2 U RNase T2 (Sigma) at 37 °C for 10 min. Reactions were quenched by the addition of 80 μl cold buffer and were immediately extracted with phenol-chloroform and precipitated with ethanol. Radiolabelled RNA products were dissolved and analysed by PAGE.

Equilibrium dialysis. The assay was performed as described in ref. 1 using 5 kDa DispoEquilibrium DIALYZERS (Harvard Apparatus). In brief, 30 μl RNA (from 0.001 to 20 μM) in the in-line probing buffer was placed in chamber A of the dialyser and equilibrated for 16 h at room temperature against chamber B containing 30 μl ^3H -labelled lysine (1 nM; ~6,000 c.p.m.) in the same buffer. The amount of bound lysine was calculated by subtracting the radioactivity counts of chamber B from chamber A. The data were fitted using a bimolecular equilibrium equation, assuming that the free lysine concentration is negligible.

31. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53**, 240–255 (1997).
32. Perrakis, A., Morris, R. & Lamzin, V. S. Automated protein model building combined with iterative structure refinement. *Nature Struct. Biol.* **6**, 458–463 (1999).

LETTERS

Concurrent nucleation of 16S folding and induced fit in 30S ribosome assembly

Tadepalli Adilakshmi^{1†}, Deepti L. Bellur^{2†} & Sarah A. Woodson¹

Rapidly growing cells produce thousands of new ribosomes each minute, in a tightly regulated process that is essential to cell growth^{1,2}. How the *Escherichia coli* 16S ribosomal RNA and the 20 proteins that make up the 30S ribosomal subunit can assemble correctly in a few minutes remains a challenging problem, partly because of the lack of real-time data on the earliest stages of assembly. By providing snapshots of individual RNA and protein interactions as they emerge in real time, here we show that 30S assembly nucleates concurrently from different points along the rRNA. Time-resolved hydroxyl radical footprinting³ was used to map changes in the structure of the rRNA within 20 milliseconds after the addition of total 30S proteins. Helical junctions in each domain fold within 100 ms. In contrast, interactions surrounding the decoding site and between the 5', the central and the 3' domains require 2–200 seconds to form. Unexpectedly, nucleotides contacted by the same protein are protected at different rates, indicating that initial RNA–protein encounter complexes refold during assembly. Although early steps in assembly are linked to intrinsically stable rRNA structure, later steps correspond to regions of induced fit between the proteins and the rRNA.

It has been previously shown that hierarchical addition of ribosomal proteins to the 16S rRNA produces cooperativity⁴ that is due to protein induced structural changes in the 16S rRNA, rather than to direct contacts between proteins⁵. Because the rRNA becomes more structured as proteins join the complex, assembly is coupled to the folding pathway of the rRNA^{6,7}. The simplest kinetic model for 30S assembly is that regions of the 16S rRNA contacted by primary assembly proteins fold first, whereas helices stabilized by tertiary assembly proteins fold last. If assembly is strictly sequential, each subdomain of the rRNA will fold within a distinct time, producing a limited set of intermediate complexes.

Alternatively, the rate of protein binding may initially depend on the stochastic probability of forming locally stable rRNA and protein interactions, with progression to later intermediates depending on propagation of this conformational order to neighbouring regions in the rRNA. If more than one region of the naked rRNA can fold, assembly is expected to nucleate from many places at once, producing an ensemble of reconstitution intermediates and multi-stage assembly kinetics⁸.

To visualize the intermediates of 30S ribosome assembly, the structure of the 16S rRNA was probed by time-resolved X-ray hydroxyl radical footprinting (Fig. 1). The extent of RNA cleavage correlates with backbone exposure⁹. Thus, this method probes individual tertiary contacts in the rRNA as well as protein interactions that bury the rRNA backbone. Previous efforts to map the conformational changes in the 16S rRNA during 30S ribosome assembly used low temperature or subsets of proteins to stall assembly at specific stages^{10,11}. We took advantage of the exceptional time resolution of synchrotron

X-ray footprinting (~10 ms) to resolve the very early stages of assembly in real time, without needing to stall the reaction.

At the start of the experiment, native 16S rRNA from *E. coli* was prefolded at 42 °C in standard reconstitution buffer (see Methods; Fig. 1a) because this pretreatment improved the quality of reconstitution (Supplementary Fig. 2). Assembly of 30S ribosomal subunits was initiated at 30 °C by mixing the rRNA with native total 30S proteins (TP30) using a rapid quench apparatus (see Supplementary Methods). The rRNA–protein mixture was exposed to a white synchrotron X-ray beam for 10 ms to generate hydroxyl radicals and to cleave the exposed parts of the rRNA backbone³. The extent of cleavage at each position in the 16S rRNA was analysed by primer extension and compared to parallel reactions on naked 16S rRNA and native 30S ribosomes (Fig. 1b).

Many regions of the prefolded 16S rRNA were strongly protected from cleavage, indicating that the rRNA has native tertiary structure in the absence of proteins. This is consistent with chemical footprinting experiments^{5,10,12} and with neutron scattering data showing that the deproteinized 16S rRNA has roughly the same dimensions as it does in the 30S ribosome¹³.

After a 3 min incubation with 30S proteins, the cleavage pattern was similar to that of native 30S ribosomes (Fig. 1b), and correlated well with the solvent accessibility of the C4' atoms predicted from crystal structures of the *Thermus thermophilus* 30S (ref. 14) or *E. coli* 70S ribosomes¹⁵. Reconstitution of 30S subunits was confirmed by sedimentation velocity and activity assays (Supplementary Fig. 2). Further rearrangements after 3 min may be needed to produce fully active subunits, as the temperature-dependent transition from the RI reconstitution intermediate to the RI* activated reconstitution intermediate late in assembly is slow at 30 °C (ref. 16), and is associated with changes in the accessibility of the nucleotide bases rather than the backbone¹¹. Nonetheless, most of the expected RNA and RNA–protein contacts formed within 3 min.

The kinetics of rRNA backbone protection showed the presence of many early assembly intermediates. Some nucleotides in the 16S rRNA were completely protected from hydroxyl radical cleavage within the 20 ms dead time of our experiment, whereas others required 1–3 min to be fully protected (Fig. 1c and Supplementary Table 1). For many nucleotides, a partial burst of protection in the first 50–100 ms was followed by slower saturation of the contact over the next few minutes (Fig. 1c and Supplementary Fig. 3).

The multiphasic folding kinetics strongly indicate that individual 30S complexes take different routes to the final structure, as suggested by previously reported rates of protein binding¹⁷ and by the kinetic partitioning of simple ribozymes among parallel folding pathways¹⁸. In kinetic footprinting experiments on the *Tetrahymena* ribozyme, partial protection of the RNA backbone at intermediate times

¹T.C. Jenkins Department of Biophysics, and ²Program in Cell, Molecular and Developmental Biology and Biophysics, Johns Hopkins University, 3400 North Charles Street, Baltimore, Maryland 21218-2685, USA. [†]Present address: Weis Center for Research, Geisinger Medical Center, 100 North Academy Avenue, Danville, Pennsylvania 17822, USA (T.A.); Department of Molecular Genetics and Cell Biology, The University of Chicago, 920 East 58th Street, Chicago, Illinois 60637, USA (D.L.B.).

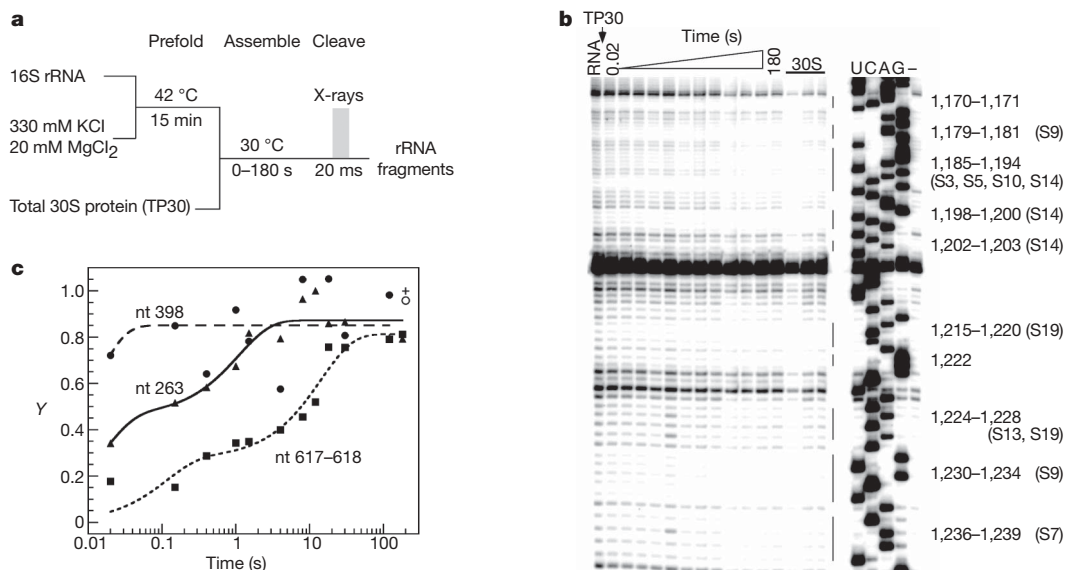


Figure 1 | Time-resolved X-ray footprinting of 30S ribosome assembly.

a, Native 16S rRNA and total protein from 30S subunits (TP30) were mixed within 5–10 ms and irradiated with a synchrotron X-ray beam to cleave the RNA at exposed ribosomes. **b**, 16S fragments were analysed by primer extension. The cleavage pattern shown was from 0.02 to 180 s after TP30 addition. 30S, triplicate controls on native 30S subunits; RNA, prefolded 16S rRNA; UCAG, dideoxy sequence ladders; –, untreated RNA. The primer

annealed after nucleotide 1,257. **c**, The relative saturation (Y) of each protection versus assembly time, fitted to single or double exponential rate equations (Supplementary Methods). Filled circles, nucleotide (nt) 398; filled squares, nucleotides 617–618; filled triangles, nucleotide 263; open circle and plus symbol, the average of 30S controls. Further data are shown in Supplementary Fig. 3 and Supplementary Table 1.

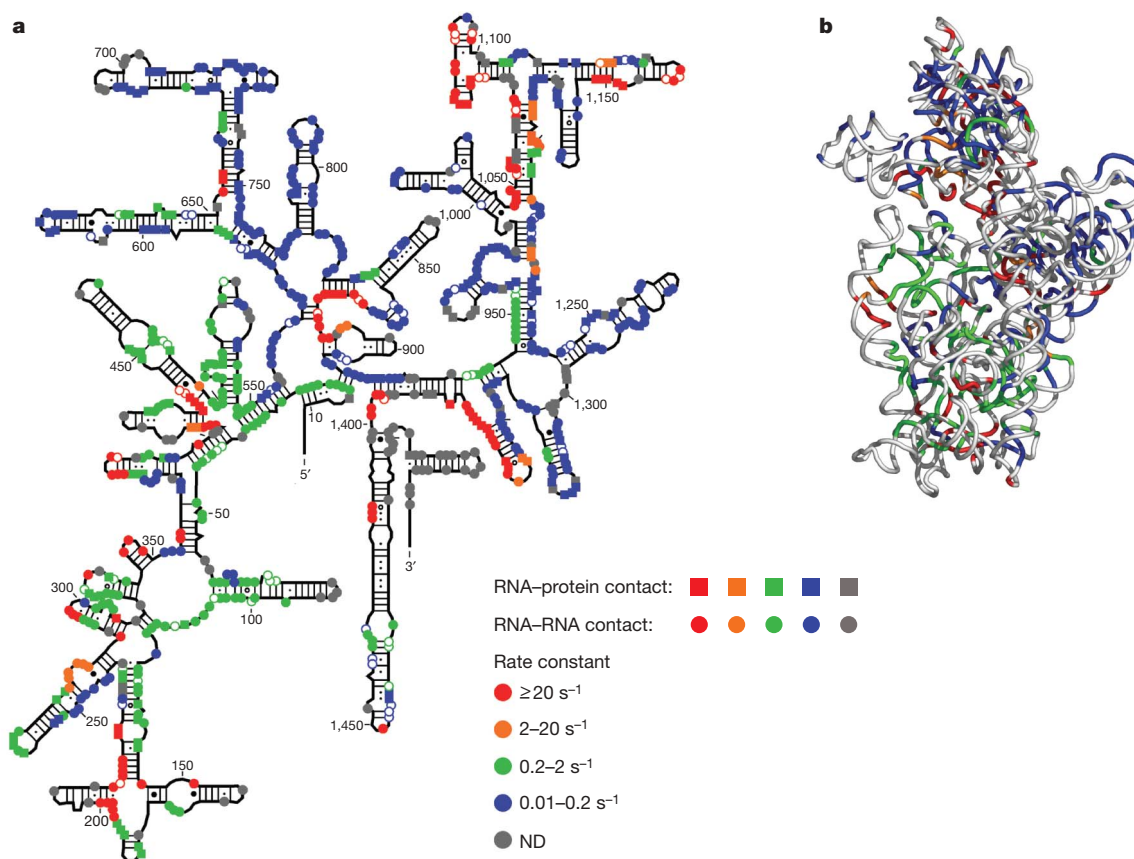


Figure 2 | Simultaneous folding of 16S domains. **a**, Protected nucleotides (740 positions) were clustered according to the rate constant for backbone protection and coloured as indicated in the key. Where the amplitude of the initial burst phase is $\leq 60\%$, the slower rate constant is used. Grey (120 positions) indicates the rate constant was undetermined (ND) owing to a pause in reverse transcription or to weak protection. Natural adenine

methylation hampered quantitative analysis of residues in the decoding site. Circles, RNA–RNA contact; squares, RNA–protein contact; open circles, predicted C4' accessible surface area $> 4 \text{ \AA}^2$; solid symbols, buried C4' in crystal structures. **b**, Three-dimensional ribbon of *E. coli* 16S rRNA (PDB accession 2avy)¹⁵ coloured as in **a**, viewed from the 50S interface. See Supplementary Fig. 4 for further views of each domain.

was explained by the contemporaneous formation of folding intermediates with different structures¹⁹. We observed no exposure of 16S residues that might indicate the disappearance of non-native assembly intermediates over time. However, such interactions may have been too dynamic or heterogeneous to produce a distinct footprint.

To visualize the assembly pathway of the rRNA, the observed rate constants for making individual backbone contacts were clustered, and the clusters were then projected onto the secondary and tertiary structure of the mature 16S rRNA (Fig. 2). This locates the interactions formed at each stage, even though the intermediate structures may differ from the mature structure. Nucleotides with similar rates of backbone protection did not map to single domains, but were distributed throughout the 16S rRNA. In general, nucleotides that were protected in the first 20–50 ms (red; Fig. 2) were partially folded in the naked rRNA or were protected by a local structure such as a kink or a helical junction. Helices contacted by the primary assembly proteins S4 (5' domain), S7 (3' domain), S8 and S15 (central domain) were also protected within 50 to 100 ms after the proteins and rRNA were mixed. Thus, each domain assembles independently and simultaneously.

In contrast, the messenger RNA decoding site and long-range interactions between domains required the longest time to form (blue; Fig. 2). For example, 16S helix H21 extends from the central domain to wrap around the body of the 30S subunit, where it contacts proteins S4 and S16. Interactions with H21 required 1 min to saturate completely ($0.02\text{--}0.1\text{ s}^{-1}$). The slow appearance of rRNA backbone contacts at G530 (H18), and the central pseudoknot (H2) in the decoding site, is consistent with previous equilibrium studies demonstrating that structural changes in the central pseudoknot are linked to the transition from the RI to RI* intermediates late in 30S assembly¹¹.

The interplay between rRNA interactions that are intrinsically stable and those that are protein-dependent is illustrated by the kinetics of RNA and RNA–protein interactions in the body of the 30S ribosome (Fig. 3a). H44 (3' minor domain) lies along one face of the body, packing against the stable tertiary structure formed by H7–H10 and H14 (ref. 20). Interactions with the distal tip of H44 are mediated by protein S20, which inserts its carboxy-terminal α -helix behind H8 and H14 in the 5' domain²¹. Many RNA tertiary interactions in the body of the 30S ribosome form within 20 ms (red; Fig. 3a), whereas interactions with S20 form in 0.5 to 3 s (green), and long-range contacts with H44 form in ~ 10 s (blue; Fig. 3a). The multi-stage assembly of these RNA and protein interactions is consistent with the results of hydroxyl radical probing from Fe-EDTA complexes tethered to S20 (ref. 22).

We next addressed whether the rates of RNA–protein interactions correlated with the position of each protein in the assembly map. Protections arising from direct protein–rRNA contacts were identified from crystallographic structures of the 30S ribosome²¹ and previous footprinting of individual proteins²³ (see Supplementary Methods). Four of the six primary assembly proteins (S4, S7, S8 and S15) protected a segment of their binding site during the first 50 ms of assembly. This was only true of three of the nine secondary assembly proteins (S16, S9 and S10) and none of the tertiary assembly proteins (Supplementary Table 2). Thus, primary assembly proteins more frequently bind to the rRNA early in assembly, in agreement with pulse-chase measurements of protein binding kinetics¹⁷.

Unexpectedly, nucleotides contacted by a single protein were protected from hydroxyl radical cleavage with different rate constants (Fig. 4). For example, protein S4 binds a five helical junction in the 5' domain, and initiates 30S assembly together with protein S7 (ref. 24). The interactions between S4 and 16S H17 (nucleotides 436–441) saturated in 20–50 ms, whereas those in H3, H16 and H18 formed in 0.3–2 s, and those with H21 (nucleotide 620) appeared over 8–10 s (Fig. 4a). S7 binds a complex helical junction in the 16S 3' domain (Fig. 4b)²¹. H43 (nucleotides 1,369–1,377) was protected in 20–50 ms, whereas contacts with H29, H41 and the distant H37 were

protected more slowly. Thus, S4 and S7 engage their binding sites in stages rather than in one step. Similar behaviour was observed for proteins S8, S9, S10, S15 and S16 (Supplementary Figs 4 and 5).

The unexpectedly complex kinetics of the rRNA–protein interactions can be explained by the formation of encounter complexes that slowly reorganize into the final complex. For S4, the variable kinetics of rRNA backbone protection may be due to co-folding of the protein and the rRNA²⁵. H17, which is protected rapidly, contacts the well-folded C-terminal domain of S4 (ref. 21). In contrast, H18 and the top of H16, which are protected at an intermediate rate, contact the S4 amino-terminal domain, which is disordered in solution²⁶.

If different regions of the 16S rRNA fold and interact with the ribosomal proteins simultaneously, the implications for the hierarchy of protein association represented by the assembly map remain a challenge. The association rates suggested by the fastest rates of RNA backbone protection approach the diffusion-controlled limit ($10^8\text{ M}^{-1}\text{ s}^{-1}$). However, the slowest protection rates are most similar to the protein binding rates measured by pulse-chase mass

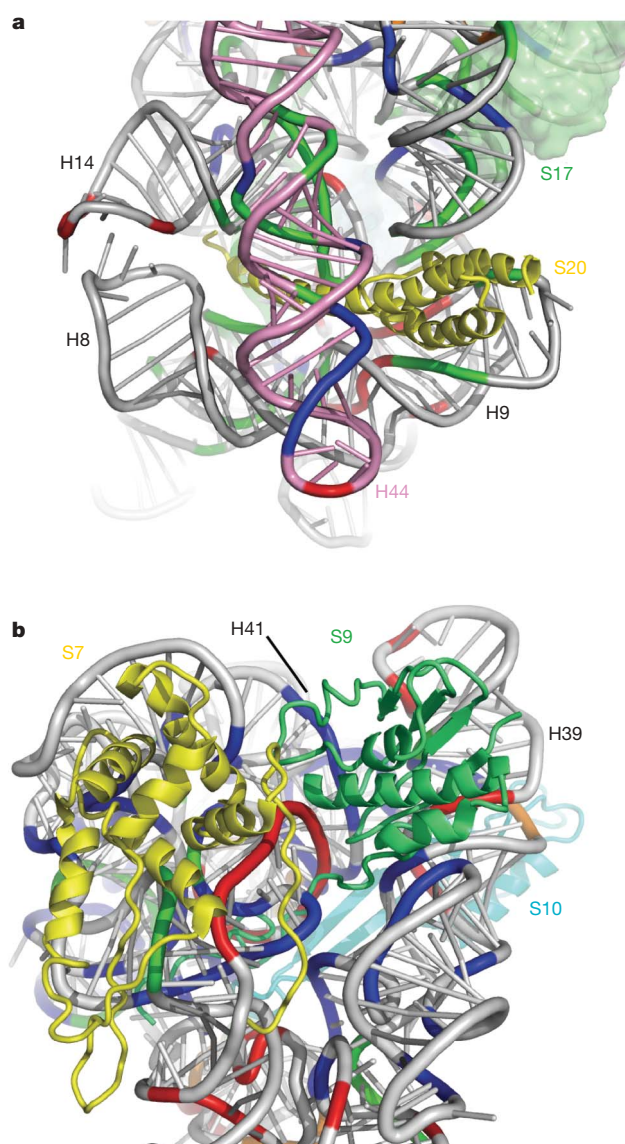


Figure 3 | Stepwise assembly of RNA and protein interactions. **a**, Protein S20 (yellow ribbon) contacts the 30S body in the 5' domain (grey) earlier than helix H44 in the 3' minor domain (pink). 16S nucleotides are coloured as in Fig. 2. **b**, Proteins S7 (yellow) and S9 (green) protect a segment of their binding site immediately (red), whereas nucleotides at the interface between the subdomains are protected slowly (blue).

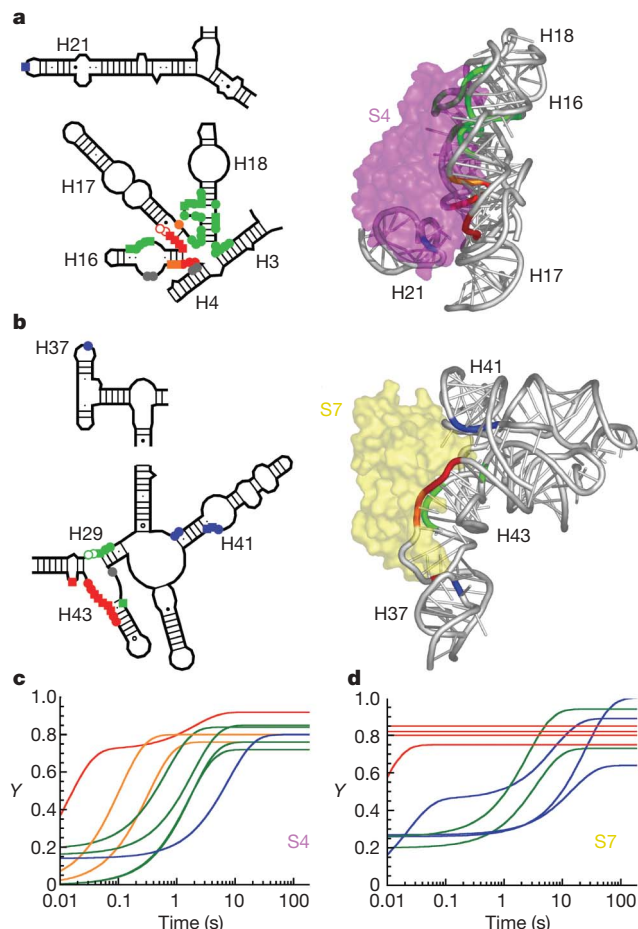


Figure 4 | Ribosomal proteins interact with the rRNA in stages.

a, b, Kinetics of direct rRNA backbone protection by proteins S4 (magenta) and S7 (yellow), coloured as in Fig. 2. Schematic symbols are also as in Fig. 2. **c, d,** Progress curves for protection of individual residues in contact with S4 and S7; for clarity, only fitted curves are shown (see data in Supplementary Fig. 5). The relative saturation (Y) of each protection versus assembly time is shown. See Supplementary Methods for definition of RNA–protein contacts.

spectrometry¹⁷ (Supplementary Table 2). Our data indicate that slow forming interactions in the rRNA, many of which are indirectly stabilized by bound proteins, determine the rate of protein addition and the hierarchy of assembly.

This idea is illustrated by interactions of protein S7, which is necessary for stable binding of other proteins to the 16S 3' domain^{24,27}. Residues indirectly protected by S7 (ref. 23), which overlap the binding sites of the secondary assembly proteins S9, S13 and S19, fold slowly (Supplementary Fig. 5), explaining why these proteins join 30S complexes more slowly than S7 (ref. 17). However, residues in H34 that directly contact S9 and S10 are fully protected in 50 ms (Fig. 3b and Supplementary Fig. 5). Thus, S7, S9 and S10 all engage their binding sites in the rRNA early in assembly, but productive incorporation of S9 and S10 presumably waits for conformational changes in the 16S rRNA induced by S7. An important question is whether early stochastic interactions between the ribosomal proteins and the rRNA backbone bias the ensemble of rRNA conformations towards the native state, thus accelerating self-assembly, or whether these interactions delay the search for the native conformation.

Stable structure in the 16S rRNA allows the concurrent nucleation of assembly from many points along the rRNA, resulting in the seemingly chaotic but rapid appearance of native interactions throughout the complex. Previous studies indicated that ribosome assembly is not completely cooperative²⁸, demonstrating the need for several nucleation sites². The lack of complete cooperativity, and the differences

between the time-dependence of 16S folding and the assembly map revealed by kinetic footprinting of the nucleotide bases¹⁰, support the conclusion that assembly proceeds in parallel through intermediates with different subsets of proteins^{6,8}. *In vivo*, assembly of the pre-rRNA during transcription is more cooperative and involves specific accessibility factors^{6,8}. Co-transcriptional assembly may simplify the pathway by limiting the number of intermediates that can be populated. Nonetheless, we expect stochastic fluctuations among alternative intermediates to contribute to the early pathway of ribosome assembly *in vivo*.

METHODS SUMMARY

Native *E. coli* 16S rRNA was prefolded in 80 mM K-cacodylate, pH 7.5, 330 mM KCl, 20 mM MgCl₂ for 15 min at 42 °C, then mixed with an equal volume of *E. coli* TP30 in the same buffer at 30 °C, using a Kin-Tek rapid quench apparatus fitted with an X-ray flow cell. Between 20% and 40% of the input RNA was cleaved after 10 or 20 ms exposure to a white light synchrotron beam (X28C, National Synchrotron Light Source (NSLS), Brookhaven National Laboratory). The cleaved RNA was analysed by primer extension with reverse transcriptase. The increase in relative backbone protection (Y) after the addition of TP30 was fitted to rate equations to obtain the observed rate constants and amplitudes for the reaction (see Supplementary Fig. 3 and Table 1). Individual protections were assigned to specific RNA or RNA–protein interactions on the basis of crystallographic structures (Protein Data Bank accessions 1j5e (ref. 14), 2avw and 2avy (ref. 15)); see Supplementary Methods for details.

Received 26 March; accepted 30 July 2008.

Published online 10 September 2008.

- Warner, J. R., Vilardell, J. & Sohn, J. H. Economics of ribosome biosynthesis. *Cold Spring Harb. Symp. Quant. Biol.* **66**, 567–574 (2001).
- Nierhaus, K. H. The assembly of prokaryotic ribosomes. *Biochimie* **73**, 739–755 (1991).
- Ralston, C. Y. et al. Time-resolved synchrotron X-ray footprinting and its application to RNA folding. *Methods Enzymol.* **317**, 353–368 (2000).
- Held, W. A., Mizushima, S. & Nomura, M. Reconstitution of *Escherichia coli* 30 S ribosomal subunits from purified molecular components. *J. Biol. Chem.* **248**, 5720–5730 (1973).
- Stern, S., Powers, T., Changchien, L. M. & Noller, H. F. RNA–protein interactions in 30S ribosomal subunits: folding and function of 16S rRNA. *Science* **244**, 783–790 (1989).
- Culver, G. M. Assembly of the 30S ribosomal subunit. *Biopolymers* **68**, 234–249 (2003).
- Williamson, J. R. After the ribosome structures: how are the subunits assembled? *RNA* **9**, 165–167 (2003).
- Noller, H. F. & Nomura, M. in *Escherichia Coli and Salmonella Typhimurium, Cellular and Molecular Biology* (ed. Neidhardt, F. C.) 104–125 (American Society for Microbiology, 1987).
- Tullius, T. D. & Greenbaum, J. A. Mapping nucleic acid structure by hydroxyl radical cleavage. *Curr. Opin. Chem. Biol.* **9**, 127–134 (2005).
- Powers, T., Daubresse, G. & Noller, H. F. Dynamics of *in vitro* assembly of 16 S rRNA into 30 S ribosomal subunits. *J. Mol. Biol.* **232**, 362–374 (1993).
- Holmes, K. L. & Culver, G. M. Mapping structural differences between 30S ribosomal subunit assembly intermediates. *Nature Struct. Mol. Biol.* **11**, 179–186 (2004).
- Adilakshmi, T., Ramaswamy, P. & Woodson, S. A. Protein-independent folding pathway of the 16S rRNA 5' domain. *J. Mol. Biol.* **351**, 508–519 (2005).
- Ramakrishnan, V. Distribution of protein and RNA in the 30S ribosomal subunit. *Science* **231**, 1562–1564 (1986).
- Wimberly, B. T. et al. Structure of the 30S ribosomal subunit. *Nature* **407**, 327–339 (2000).
- Schuwirth, B. S. et al. Structures of the bacterial ribosome at 3.5 Å resolution. *Science* **310**, 827–834 (2005).
- Traub, P. & Nomura, M. Structure and function of *Escherichia coli* ribosomes. VI. Mechanism of assembly of 30 S ribosomes studied *in vitro*. *J. Mol. Biol.* **40**, 391–413 (1969).
- Talkington, M. W., Siuzdak, G. & Williamson, J. R. An assembly landscape for the 30S ribosomal subunit. *Nature* **438**, 628–632 (2005).
- Pan, J., Thirumalai, D. & Woodson, S. A. Folding of RNA involves parallel pathways. *J. Mol. Biol.* **273**, 7–13 (1997).
- Laederach, A., Shcherbakova, I., Liang, M. P., Brenowitz, M. & Altman, R. B. Local kinetic measures of macromolecular structure reveal partitioning among multiple parallel pathways from the earliest steps in the folding of a large RNA molecule. *J. Mol. Biol.* **358**, 1179–1190 (2006).
- Cate, J. H., Yusupov, M. M., Yusupova, G. Z., Earnest, T. N. & Noller, H. F. X-ray crystal structures of 70S ribosome functional complexes. *Science* **285**, 2095–2104 (1999).
- Brodersen, D. E., Clemons, W. M. Jr, Carter, A. P., Wimberly, B. T. & Ramakrishnan, V. Crystal structure of the 30 S ribosomal subunit from *Thermus*

- thermophilus: structure of the proteins and their interactions with 16 S RNA. *J. Mol. Biol.* **316**, 725–768 (2002).
22. Dutca, L. M. & Culver, G. M. Assembly of the 5' and 3' minor domains of 16S ribosomal RNA as monitored by tethered probing from ribosomal protein S20. *J. Mol. Biol.* **376**, 92–108 (2008).
23. Powers, T. & Noller, H. F. Hydroxyl radical footprinting of ribosomal proteins on 16S rRNA. *RNA* **1**, 194–209 (1995).
24. Nowotny, V. & Nierhaus, K. H. Assembly of the 30S subunit from *Escherichia coli* ribosomes occurs via two assembly domains which are initiated by S4 and S7. *Biochemistry* **27**, 7051–7055 (1988).
25. Powers, T. & Noller, H. F. A temperature-dependent conformational rearrangement in the ribosomal protein S4·16 S rRNA complex. *J. Biol. Chem.* **270**, 1238–1242 (1995).
26. Sayers, E. W., Gerstner, R. B., Draper, D. E. & Torchia, D. A. Structural preordering in the N-terminal region of ribosomal protein S4 revealed by heteronuclear NMR spectroscopy. *Biochemistry* **39**, 13602–13613 (2000).
27. Samaha, R. R., O'Brien, B., O'Brien, T. W. & Noller, H. F. Independent *in vitro* assembly of a ribonucleoprotein particle containing the 3' domain of 16S rRNA. *Proc. Natl Acad. Sci. USA* **91**, 7884–7888 (1994).
28. Dodd, J., Kolb, J. M. & Nomura, M. Lack of complete cooperativity of ribosome assembly *in vitro* and its possible relevance to *in vivo* ribosome assembly and the regulation of ribosomal gene expression. *Biochimie* **73**, 757–767 (1991).
- Supplementary Information** is linked to the online version of the paper at www.nature.com/nature.
- Acknowledgements** We thank R. Moss, A. Cukras, L. Cochella and R. Green for help with ribosome preparation and peptidyl transferase assays, P. Fleming for help with Calc-Surf software, and S. Gupta, M. Sullivan and M. Brenowitz for help with X-ray footprinting. This work was supported by the National Institutes of Health (NIH; GM60819). The NSLS X28C and the Center for Synchrotron Biosciences are supported by NIH P41-EB0001979.
- Author Contributions** T.A. performed the experiments, analysed the data and prepared the figures; D.L.B. analysed protections in the 3' minor domain; and S.A.W. prepared the figures and wrote the paper.
- Author Information** Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to S.A.W. (swoodson@jhu.edu).

naturejobs

**THE CAREERS
MAGAZINE FOR
SCIENTISTS**

Much like every other industry, the biotechnology sector is bracing itself for a difficult ride in the present financial crisis. That the industry has suffered hard times before may give it an advantage, but it is still likely to find the squeeze on cash difficult, and the knock-on effects for jobs could put some regions at a disadvantage.

David Mott, a venture capitalist at New Enterprise Associates in Chevy Chase, Maryland, describes the current investment landscape for the biotech sector as somewhat desolate and dried up. Without ample investment, existing biotech companies will be harder to sustain, and new firms will be trickier to set up. As a result, it will be harder to generate jobs in the sector.

But Mott, the former chief executive of MedImmune, a biotech company bought by AstraZeneca in April 2007 for \$15.6 billion, suggests that the industry is well prepared to survive — if only because it has suffered already. The last big biotech liquidity crunch was in the early and mid 1990s. “There was very little capital available to build even very good companies with very good management teams,” says Mott. That means many people in the sector have already been through dry spells in the capital markets, he notes. “That experience will come in handy.”

Perhaps more importantly, the biotech sector has already adapted to a low-liquidity environment. Financing models for the area were forced to evolve after the technology bubble burst in 2000 and public offerings more or less dried up. Biotech investors and founders are no longer attempting to build the next Amgen or Genentech. Instead, they are building companies to sell, usually hoping big pharmaceutical firms will snap them up. This, says Mott, makes many of today’s biotech concerns ‘virtual businesses’ — they are increasingly outsourcing chemistry and clinical-trials management to places such as India.

The deepening economic crisis is likely to accentuate this trend. So it could be good news for Indian scientists who are seeking biotech-related jobs, but it may be less welcome for those in the United States and Europe.

Gene Russo is editor of *Naturejobs*.

CONTACTS

Editor: Gene Russo

European Head Office, London
The Macmillan Building,
4 Crinan Street, London N1 9XW, UK
Tel: +44 (0) 20 7843 4961
Fax: +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

European Sales Manager:
Andy Douglas (4975)
e-mail: a.douglas@nature.com

Natureevents:
Ghizlaine Ababou (+44 (0) 20 7014 4015)
e-mail: g.ababou@nature.com

UK Corporate:

Nils Moeller (4953)

Southwest UK/RoW:

Alexander Ranken (4944)

Northeast UK/Ireland:

Matthew Ward (+44 (0) 20 7014 4059)

France/Switzerland/Belgium:

Muriel Lestringuez (4994)

Scandinavia/Spain/Portugal/Italy:

Evelina Rubio-Hakansson (4973)

North Germany/The Netherlands/Eastern

Europe: Reya Silao (4970)

South Germany/Austria:

Hildi Rowland (+44 (0) 20 7014 4084)

Advertising Production Manager:

Stephen Russell

To send materials use London address above.

Tel: +44 (0) 20 7843 4816

Fax: +44 (0) 20 7843 4996

e-mail: naturejobs@nature.com

Naturejobs web development: Tom Hancock

Naturejobs online production: Dennis Chu

US Head Office, New York

75 Varick Street, 9th Floor,

New York, NY 10013-1917

Tel: +1 800 989 7718

Fax: +1 800 989 7103

e-mail: naturejobs@natureny.com

US Sales Manager: Peter Bless

India

Vikas Chawla (+91 1242881057)

e-mail: v.chawla@nature.com

Japan Head Office, Tokyo

Chiyoda Building, 2-37 Ichigayatamachi,

Shinjuku-ku, Tokyo 162-0843

Tel: +81 3 3267 8751

Fax: +81 3 3267 8746

Asia-Pacific Sales Manager:

Ayako Watanabe (+81 3 3267 8765)

e-mail: a.watanabe@natureasia.com

Business Development Manager, Greater

China/Singapore:

Gloria To (+852 2811 7191)

e-mail: g.to@natureasia.com



PHYSICAL IMPACT

Despite its diminutive size and modest research budget, the Netherlands makes a major contribution to research in the physical sciences. **Quirin Schiermeier** reports.

Twenty-nine year old American astronomer Ryan Quadri was at a point in his career when he found the idea of living abroad appealing. After earning his PhD from Yale University last year, Quadri, who studies the formation of galaxies in the early Universe, considered various destinations. He preferred Europe, and took advantage of a connection he had with a Dutch astronomer. "Living in another country was my main motivation," he says. "But I did know, of course, that Leiden is a great place for astronomy." So last November, Quadri relocated to Leiden, a lively university town an hour's train ride south of Amsterdam, to do three years of postdoctoral research. He has no regrets.

Quadri says it was surprisingly easy to get used to working and living in the Netherlands, a small place with strong scientific hubs in cities such as Leiden and Utrecht. The country has long been known for a welcoming attitude towards research and a history of encouraging scientific discourse. Scientists there have learned to forge international collaborations rather than be confined by the resources of a small country. Quadri also found that the lab culture was very similar to the one he had left behind.

The Netherlands boasts significant achievements in astronomy and other physical sciences, given its size. In August, researchers at Dutch institutions won nine out of 105 advanced grants in the physical sciences from the European Research Council (ERC), making the Netherlands the third most successful nation — behind Britain and France and on par with Italy and Switzerland — in this round of the competition. Researchers in Germany, which is five times more populous than the Netherlands, have won only seven of these prestigious new grants.

"When I moved to the Netherlands in 2001, I was really impressed by the general quality of physics in this

country," says Michel Orrit, a French condensed-matter physicist at the University of Leiden, who has won an €1.8-million (US\$2.4-million) ERC grant for his work on single-molecule spectroscopy.

And yet the Netherlands' overall spending on science, some 1.8% of gross domestic product, is far below that of its eastern neighbour and even slightly below the European Union (EU) average.

"Call it the Dutch paradox," says Peter Nijkamp, chairman of the Netherlands Organisation for Scientific Research (NWO), the main funding agency. Nijkamp suggests that the Netherlands has managed to create very favourable conditions for science despite "regrettably modest public research budgets".

A city of science

Leiden is one of the country's centres for physical sciences. With a population of 117,000, it lies in the Holland region, a highly urbanized ring in the west of country and one of the most densely populated areas of the world. An old windmill, quiet canals, flocks of cyclists and the odd coffee shop remind one that this is the land of Vincent van Gogh, cyclist Joop Zoetemelk and pragmatic cannabis policies.

But just below Leiden's tourist-pleasing surface is a rich and active tradition of science. The Leiden Bioscience Park is one of the oldest and largest in Europe. The university, where the seventeenth-century astronomer Christiaan Huygens studied law and mathematics, also hosts the country's largest and best-equipped astronomy department. Most research takes place in international collaboration, thus affording privileged access to telescopes, data and computing facilities. And thanks to a generous grant from the Netherlands Research School for Astronomy, Quadri can also travel to conferences and observatory sites around the world as often as necessary.



Jaap Sinninghe Damsté (top) and Michel Orrit have received European Research Council grants for their work.

J. HICKS/CORBIS

Indeed, astronomy is one of the country's prime science strengths. Dutch space researchers often work with international colleagues, while trying to maintain a strong domestic base for training students and young scientists. In the past 20 years, more PhD students from Leiden have been selected for NASA's prestigious Hubble postdoctoral fellowship programme than from any other university outside the United States.

"Astronomy is a very attractive field for young physicists," says Quadri's original contact Marijn Franx, a group leader at Leiden University who was awarded a Hubble fellowship in 1991. Franx notes that former students have found positions in fields as diverse as high-energy physics, software development, publishing, funding agencies and even the diplomatic service. In August, Franx received a €1.5-million ERC advanced grant for his research on how distant galaxies come into being. He plans to build up a new group comprising a tenure-track faculty member, two postdocs and two PhD students, to extend his work on high-redshift galaxies by looking even further back in time.

Critical evaluation

The key to success is minimal bureaucracy and an efficient, highly competitive science-funding system, says Nijkamp. Targeted support mechanisms, for both early-stage and senior researchers, include the NWO's Rubicon grants for postdoctoral scientists and the prestigious €1.5-million Spinoza prize for outstanding achievement. A strict system of performance measurement for scientists and university departments helps grant reviewers allocate resources efficiently.

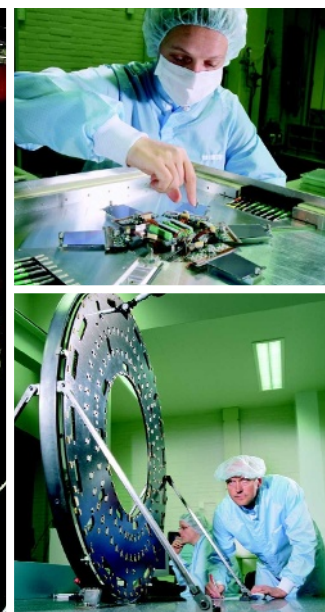
Starting in the 1980s, Dutch universities developed a research and teaching evaluation system involving external reviewers. Criteria include publications and success in attracting grant money. Those who measure up have a much better shot at tenure. Nijkamp says the system has been a big help in improving the quality of Dutch science, although there are winners and losers. "Our scientists have learned that they will be strictly evaluated, and that their success in getting funded depends only on their performance," says Nijkamp.

All NWO programmes are heavily oversubscribed, with grant success rates in some fields as low as 10%. The tough competition, says Nijkamp, does create bottlenecks. "We often have to say 'no' just because the money isn't there," he says. "That's a pity, because it means we're missing out on too many good people."

But tight national research budgets have not diminished the Netherlands' attractiveness as a host for young scientists from abroad. Terms of employment for PhD students are particularly appealing. They are hired on a full-time basis for four years — rather than for three years as in most other European countries, where they often have only student status. And they get paid unusually well: in their fourth year, they can expect a salary of €3,000 a month before tax.

The extra year gives PhD students more time to transform their research into papers, says Jaap Sinninghe Damsté, a molecular palaeontologist at the Royal Netherlands Institute for Sea Research (NIOZ) on the Friesian island of Texel. A good publication record early on helps those who want to stay in research find attractive postdoctoral positions, he says.

Not many young scientists have luck founding new academic labs, however. Sinninghe Damsté, who has won an ERC advanced grant for his method of using



Dutch involvement in physics: (clockwise from left) helping to build the TEXTOR plasma reactor, equipment at the National Institute for Subatomic Physics, and working on a component for the Large Hadron Collider.

fossil molecules in sediments to reconstruct past climate, is a rare example of a successful Dutch scientist who has never worked outside his native country. With the €2.5-million grant he will build up a new group in organic biogeochemistry, comprising three PhD students, two postdocs and a technician. For the first time in many years, the NIOZ — one of ten research institutes run by the NWO — will also create two or three new tenure-track positions.

Generally, however, a young scientist's prospects of finding permanent positions there are bleak. "To try to become a professional scientist here you really have to love research very much," says Sinninghe Damsté.

"It's hard," agrees his postdoc Jung-Hyun Kim. A marine geologist from South Korea, Kim did a postdoc in Bremen, Germany, before moving on to the NIOZ. She came to the Netherlands on an EU-funded Marie Curie fellowship. Despite the challenges, she says, the combination of an open, multicultural society and high standards in science make the country an ideal training ground for young researchers.

Like some other countries, such as Germany, most science jobs in the Netherlands are occupied by postdocs on fixed-term contracts. Kim is therefore putting out feelers towards France, where her husband lives and where she hopes it will be easier to find a permanent position. Her ideal is an ERC starting grant that would help her to build up an independent research team of her own.

When his contract in Leiden expires, Quadri, too, plans to leave the Netherlands and head back to the United States. Despite its success in attracting foreign talent, the Netherlands often remains a stopping-off place for young scientists early in their careers.

Even so, Dutch science as a whole benefits from the flow of international talent — and those who go there benefit from a vibrant and welcoming research community. "I got so into astronomy here," says Quadri, "that all other alternatives disappeared."

Quirin Schiermeier is Nature's Germany correspondent.



Jung-Hyun Kim (top) and Ryan Quadri have both enjoyed postdocs in the Netherlands.

MOVERS

Colin Lonsdale, director, Haystack Observatory, Massachusetts Institute of Technology, Westford, Massachusetts



2006–08: Assistant director, Haystack Observatory, Massachusetts Institute of Technology

1986–2008: Research scientist, Haystack Observatory, Massachusetts Institute of Technology

1983–86: Research associate, Pennsylvania State University, University Park, Pennsylvania

Colin Lonsdale says radio astronomy is experiencing a revolution. The new director of the Haystack Observatory in Westford, part of the Massachusetts Institute of Technology (MIT), believes that new instruments will probe unexplored parts of the distant Universe and allow scientists to answer crucial questions about how stars and galaxies first formed and continue to grow.

This is the second such revolution for Lonsdale. When working on his PhD in the early 1980s, he demonstrated how energy is transported in extragalactic radio sources by studying radio jets and hotspots, using the then-new MERLIN array run from Jodrell Bank Observatory at the University of Manchester, UK. He calls his decision to pursue a PhD there “career-defining”.

As a postdoc at Pennsylvania State University, Lonsdale used interferometry techniques to study all aspects of extragalactic radio sources. His expertise in data-reduction software led him to a research scientist position at the Haystack Observatory. “That’s when things took off for me,” says Lonsdale, who helped to detect an unusual radio signal from a complex, ultraluminous infrared galaxy created by the explosions of 50 supernovae.

That discovery, made together with his astronomer sister Carol, stunned colleagues. It not only yielded the first images of many simultaneously exploding stars inside a galaxy, but also altered the prevailing theory of how galaxies generate intense radio sources called masers, according to long-time colleague Phil Diamond, director of the Jodrell Bank Centre for Astrophysics. Diamond says it took Lonsdale’s imagination and tenacity to demonstrate that these galaxies are starbursts.

In 2000, Lonsdale turned away from pure research, instead devoting his time to developing the Murchison Widefield Array (MWA), an array of low-frequency radio telescopes in Western Australia. He says the MWA, scheduled to begin operation in 2010, will target questions such as how to measure solar storms accurately.

Lonsdale says advances in digital electronics have opened up a new frontier in radio science. Diamond fully expects that Lonsdale, who got funding for the MWA in difficult times, will use his expertise to help position MIT’s involvement in the next generation of radio-telescope technology — the Square Kilometre Array. Diamond is optimistic about the effect Lonsdale can have on this project, which he calls “the big daddy of radio astronomy”. ■

Virginia Gewin

NETWORKS & SUPPORT

Sustenance for sustainability

Scientists who seek intensely interdisciplinary study could be the beneficiaries of increasing interest in the emerging field of sustainability research, with new university programmes offering novel opportunities. Portland State University in Oregon and Cornell University in Ithaca, New York, are the most recent entrants to the field. They follow the example set by institutions such as the School of Sustainability at Arizona State University in Tempe.

Broadly defined, sustainability bridges disciplines to determine how to meet the resource needs of the present without adversely affecting future generations. In practice, that means assembling teams of ecologists, economists, biologists and social scientists to find solutions and strategies for big problems, such as meeting future energy needs.

Portland has received US\$25 million from the James F. and Marion L. Miller Foundation towards sustainability research and teaching, and Cornell’s \$10-million National Science Foundation (NSF) grant will be used to develop the field of ‘computational sustainability’ — using computer-science techniques to manage natural resources more effectively.

Portland’s provost, Roy Koch, says their approach is to connect

curiosity-driven research to short-term goals. Within the next 2 years, they plan to recruit at least 10 faculty members, 10–20 graduate students and up to 5 postdocs to tackle technical issues surrounding alternative energy sources, sustainable urban communities and developing the metrics of sustainability. He also hopes to set up a visiting faculty programme to forge national and international connections.

Cornell’s Institute for Computational Sustainability involves scientists from both Cornell and other US research institutions, whose work will involve merging disparate data sets related to areas such as environmental science and economics. One example is modelling the evolution of fish populations to assist with the designation of no-fishing zones.

The main focus of the NSF grant is to train some 12 graduate students each year. Cornell also wants to attract postdocs who are interested in computer science or applied maths. Carla Gomes, the institute’s director, says that “computer scientists aren’t aware that their expertise can impact issues of sustainability, and we hope to inject computational thinking into problem solving.” ■

Virginia Gewin

POSTDOC JOURNAL

Waiting with baboons

I have found it hard to return to my research in Ethiopia after my six-week hiatus. The fear of the unknown is gone, but so is some of the excitement. And this time, I have no one from the outside world with me. The scouts and our assistant are around, but there is nobody with whom to muse about science over a warmish beer. And, as any semi-sane person can tell you, having too much time for your own thoughts can be utterly destructive.

For the first few days here, I ranted and raved silently, even finding it difficult to go outdoors. It was too much ‘reality’ hitting me — I’m here and I’m more or less alone. I considered the gelada baboons my only distraction, and negativity threatened to devour me. I had to try to see the positive aspects of this experience. After all, I want to be here and I love my job. How can I deal with the circumstances surrounding it?

I came to see that I have been given something that I never had in the city: time. It is a precious thing, and not all that frightening. Now I am filling my time with the things I never get to do back home — reading articles, doing extra analyses, working overtime and writing letters. And I’m finding it a positive experience. I still love all social distractions, but I don’t constantly bemoan my solitude. Only my attitude has changed. Perhaps that’s enough to carry me through. ■

Aliza le Roux is a postdoctoral fellow in animal behaviour at the University of Michigan.

MOVERS

Colin Lonsdale, director, Haystack Observatory, Massachusetts Institute of Technology, Westford, Massachusetts



2006–08: Assistant director, Haystack Observatory, Massachusetts Institute of Technology

1986–2008: Research scientist, Haystack Observatory, Massachusetts Institute of Technology

1983–86: Research associate, Pennsylvania State University, University Park, Pennsylvania

Colin Lonsdale says radio astronomy is experiencing a revolution. The new director of the Haystack Observatory in Westford, part of the Massachusetts Institute of Technology (MIT), believes that new instruments will probe unexplored parts of the distant Universe and allow scientists to answer crucial questions about how stars and galaxies first formed and continue to grow.

This is the second such revolution for Lonsdale. When working on his PhD in the early 1980s, he demonstrated how energy is transported in extragalactic radio sources by studying radio jets and hotspots, using the then-new MERLIN array run from Jodrell Bank Observatory at the University of Manchester, UK. He calls his decision to pursue a PhD there “career-defining”.

As a postdoc at Pennsylvania State University, Lonsdale used interferometry techniques to study all aspects of extragalactic radio sources. His expertise in data-reduction software led him to a research scientist position at the Haystack Observatory. “That’s when things took off for me,” says Lonsdale, who helped to detect an unusual radio signal from a complex, ultraluminous infrared galaxy created by the explosions of 50 supernovae.

That discovery, made together with his astronomer sister Carol, stunned colleagues. It not only yielded the first images of many simultaneously exploding stars inside a galaxy, but also altered the prevailing theory of how galaxies generate intense radio sources called masers, according to long-time colleague Phil Diamond, director of the Jodrell Bank Centre for Astrophysics. Diamond says it took Lonsdale’s imagination and tenacity to demonstrate that these galaxies are starbursts.

In 2000, Lonsdale turned away from pure research, instead devoting his time to developing the Murchison Widefield Array (MWA), an array of low-frequency radio telescopes in Western Australia. He says the MWA, scheduled to begin operation in 2010, will target questions such as how to measure solar storms accurately.

Lonsdale says advances in digital electronics have opened up a new frontier in radio science. Diamond fully expects that Lonsdale, who got funding for the MWA in difficult times, will use his expertise to help position MIT’s involvement in the next generation of radio-telescope technology — the Square Kilometre Array. Diamond is optimistic about the effect Lonsdale can have on this project, which he calls “the big daddy of radio astronomy”. ■

Virginia Gewin

NETWORKS & SUPPORT

Sustenance for sustainability

Scientists who seek intensely interdisciplinary study could be the beneficiaries of increasing interest in the emerging field of sustainability research, with new university programmes offering novel opportunities. Portland State University in Oregon and Cornell University in Ithaca, New York, are the most recent entrants to the field. They follow the example set by institutions such as the School of Sustainability at Arizona State University in Tempe.

Broadly defined, sustainability bridges disciplines to determine how to meet the resource needs of the present without adversely affecting future generations. In practice, that means assembling teams of ecologists, economists, biologists and social scientists to find solutions and strategies for big problems, such as meeting future energy needs.

Portland has received US\$25 million from the James F. and Marion L. Miller Foundation towards sustainability research and teaching, and Cornell’s \$10-million National Science Foundation (NSF) grant will be used to develop the field of ‘computational sustainability’ — using computer-science techniques to manage natural resources more effectively.

Portland’s provost, Roy Koch, says their approach is to connect

curiosity-driven research to short-term goals. Within the next 2 years, they plan to recruit at least 10 faculty members, 10–20 graduate students and up to 5 postdocs to tackle technical issues surrounding alternative energy sources, sustainable urban communities and developing the metrics of sustainability. He also hopes to set up a visiting faculty programme to forge national and international connections.

Cornell’s Institute for Computational Sustainability involves scientists from both Cornell and other US research institutions, whose work will involve merging disparate data sets related to areas such as environmental science and economics. One example is modelling the evolution of fish populations to assist with the designation of no-fishing zones.

The main focus of the NSF grant is to train some 12 graduate students each year. Cornell also wants to attract postdocs who are interested in computer science or applied maths. Carla Gomes, the institute’s director, says that “computer scientists aren’t aware that their expertise can impact issues of sustainability, and we hope to inject computational thinking into problem solving.” ■

Virginia Gewin

POSTDOC JOURNAL

Waiting with baboons

I have found it hard to return to my research in Ethiopia after my six-week hiatus. The fear of the unknown is gone, but so is some of the excitement. And this time, I have no one from the outside world with me. The scouts and our assistant are around, but there is nobody with whom to muse about science over a warmish beer. And, as any semi-sane person can tell you, having too much time for your own thoughts can be utterly destructive.

For the first few days here, I ranted and raved silently, even finding it difficult to go outdoors. It was too much ‘reality’ hitting me — I’m here and I’m more or less alone. I considered the gelada baboons my only distraction, and negativity threatened to devour me. I had to try to see the positive aspects of this experience. After all, I want to be here and I love my job. How can I deal with the circumstances surrounding it?

I came to see that I have been given something that I never had in the city: time. It is a precious thing, and not all that frightening. Now I am filling my time with the things I never get to do back home — reading articles, doing extra analyses, working overtime and writing letters. And I’m finding it a positive experience. I still love all social distractions, but I don’t constantly bemoan my solitude. Only my attitude has changed. Perhaps that’s enough to carry me through. ■

Aliza le Roux is a postdoctoral fellow in animal behaviour at the University of Michigan.

MOVERS

Colin Lonsdale, director, Haystack Observatory, Massachusetts Institute of Technology, Westford, Massachusetts



2006–08: Assistant director, Haystack Observatory, Massachusetts Institute of Technology

1986–2008: Research scientist, Haystack Observatory, Massachusetts Institute of Technology

1983–86: Research associate, Pennsylvania State University, University Park, Pennsylvania

Colin Lonsdale says radio astronomy is experiencing a revolution. The new director of the Haystack Observatory in Westford, part of the Massachusetts Institute of Technology (MIT), believes that new instruments will probe unexplored parts of the distant Universe and allow scientists to answer crucial questions about how stars and galaxies first formed and continue to grow.

This is the second such revolution for Lonsdale. When working on his PhD in the early 1980s, he demonstrated how energy is transported in extragalactic radio sources by studying radio jets and hotspots, using the then-new MERLIN array run from Jodrell Bank Observatory at the University of Manchester, UK. He calls his decision to pursue a PhD there “career-defining”.

As a postdoc at Pennsylvania State University, Lonsdale used interferometry techniques to study all aspects of extragalactic radio sources. His expertise in data-reduction software led him to a research scientist position at the Haystack Observatory. “That’s when things took off for me,” says Lonsdale, who helped to detect an unusual radio signal from a complex, ultraluminous infrared galaxy created by the explosions of 50 supernovae.

That discovery, made together with his astronomer sister Carol, stunned colleagues. It not only yielded the first images of many simultaneously exploding stars inside a galaxy, but also altered the prevailing theory of how galaxies generate intense radio sources called masers, according to long-time colleague Phil Diamond, director of the Jodrell Bank Centre for Astrophysics. Diamond says it took Lonsdale’s imagination and tenacity to demonstrate that these galaxies are starbursts.

In 2000, Lonsdale turned away from pure research, instead devoting his time to developing the Murchison Widefield Array (MWA), an array of low-frequency radio telescopes in Western Australia. He says the MWA, scheduled to begin operation in 2010, will target questions such as how to measure solar storms accurately.

Lonsdale says advances in digital electronics have opened up a new frontier in radio science. Diamond fully expects that Lonsdale, who got funding for the MWA in difficult times, will use his expertise to help position MIT’s involvement in the next generation of radio-telescope technology — the Square Kilometre Array. Diamond is optimistic about the effect Lonsdale can have on this project, which he calls “the big daddy of radio astronomy”. ■

Virginia Gewin

NETWORKS & SUPPORT

Sustenance for sustainability

Scientists who seek intensely interdisciplinary study could be the beneficiaries of increasing interest in the emerging field of sustainability research, with new university programmes offering novel opportunities. Portland State University in Oregon and Cornell University in Ithaca, New York, are the most recent entrants to the field. They follow the example set by institutions such as the School of Sustainability at Arizona State University in Tempe.

Broadly defined, sustainability bridges disciplines to determine how to meet the resource needs of the present without adversely affecting future generations. In practice, that means assembling teams of ecologists, economists, biologists and social scientists to find solutions and strategies for big problems, such as meeting future energy needs.

Portland has received US\$25 million from the James F. and Marion L. Miller Foundation towards sustainability research and teaching, and Cornell’s \$10-million National Science Foundation (NSF) grant will be used to develop the field of ‘computational sustainability’ — using computer-science techniques to manage natural resources more effectively.

Portland’s provost, Roy Koch, says their approach is to connect

curiosity-driven research to short-term goals. Within the next 2 years, they plan to recruit at least 10 faculty members, 10–20 graduate students and up to 5 postdocs to tackle technical issues surrounding alternative energy sources, sustainable urban communities and developing the metrics of sustainability. He also hopes to set up a visiting faculty programme to forge national and international connections.

Cornell’s Institute for Computational Sustainability involves scientists from both Cornell and other US research institutions, whose work will involve merging disparate data sets related to areas such as environmental science and economics. One example is modelling the evolution of fish populations to assist with the designation of no-fishing zones.

The main focus of the NSF grant is to train some 12 graduate students each year. Cornell also wants to attract postdocs who are interested in computer science or applied maths. Carla Gomes, the institute’s director, says that “computer scientists aren’t aware that their expertise can impact issues of sustainability, and we hope to inject computational thinking into problem solving.” ■

Virginia Gewin

POSTDOC JOURNAL

Waiting with baboons

I have found it hard to return to my research in Ethiopia after my six-week hiatus. The fear of the unknown is gone, but so is some of the excitement. And this time, I have no one from the outside world with me. The scouts and our assistant are around, but there is nobody with whom to muse about science over a warmish beer. And, as any semi-sane person can tell you, having too much time for your own thoughts can be utterly destructive.

For the first few days here, I ranted and raved silently, even finding it difficult to go outdoors. It was too much ‘reality’ hitting me — I’m here and I’m more or less alone. I considered the gelada baboons my only distraction, and negativity threatened to devour me. I had to try to see the positive aspects of this experience. After all, I want to be here and I love my job. How can I deal with the circumstances surrounding it?

I came to see that I have been given something that I never had in the city: time. It is a precious thing, and not all that frightening. Now I am filling my time with the things I never get to do back home — reading articles, doing extra analyses, working overtime and writing letters. And I’m finding it a positive experience. I still love all social distractions, but I don’t constantly bemoan my solitude. Only my attitude has changed. Perhaps that’s enough to carry me through. ■

Aliza le Roux is a postdoctoral fellow in animal behaviour at the University of Michigan.

Faux-pas, Doc

Don't toy with affections.

Janett L. Grady

I'm in Paris, on my way to Boston. My insides are all screwed up and I'm on my way to see if I can get things fixed. If I can't get things fixed, the conservative types in World Congress will order me taken apart and I'll be fed to the hogs, piece by piece. I'm three hours early for the transfer and I've already been scanned and labelled. I'm killing time, wandering back and forth, when I suddenly spot her, or think it's her. She's in one of those ancient hoverchairs. I wave, catch her eye and she waves back, hovers towards me, a quizzical, surprised look on her face.

"Well, well," she says with a smile. "What a surprise."

I smile in return. "Doc," I say, "it is you. My God, you haven't changed at all." It's almost true. She's in the chair, a few pounds heavier, a little older-looking but the same Doc Tan who turned me on back in 3010. "How have you been?" I ask. "It's been — what — 20 years?"

"At least that long," she says. She hovers back a little and looks me up and down, her blue eyes crinkling at the corners. She's gorgeous. Even after all these years, I still want to give her a great big kiss. "Damn, T-3," she says, "you're looking fine." She sighs and then frowns prettily. "I never thought you'd last this long," she says. "Is it all still working?"

"Still working," I tell her, and hesitate, unsure about telling her what's wrong. "I'm at the Capitol," I tell her. "I've been there for nine years."

"Men, women or what?" she asks. "You're wired for both, or at least I tried." She laughs. "But with all the women in World Congress these days, I wouldn't be all that surprised if you've been rewired to prefer women."

"You're right," I tell her. "I still do men, but I've been tweaked to prefer women."

"I figured they might do that," she says. "But, hey, no big deal, right?"

"It is a big deal," I shoot back. "When I do men, it hurts like hell. There's something wrong. I'm on my way to Mass Tech to see if I can get it fixed."

She doesn't seem interested. She shrugs, looks at her watch. "I've got a few minutes," she says. "You want some wine?" She doesn't wait for an answer. She spins around, waves for me to follow, and leads the way into the French Space. She leaves me at a table, goes for the wine and then joins me. She sips from her glass, settles

back in her chair. "You do look a little different," she says, "but hell, don't we all? Me, I'm stuck in this chair." Her eyes seem fixed on my blouse. "So what's wrong?" she asks. "You look healthy enough."

"I'm not," I tell her. "The penis-pocket is dosing shut, not to mention being backed up. I'm not passing the way I should be."

"It's been a long time," she says. "You've been screwing your brains out, and you're scratched and swollen. It's the inside sensors. They wear out." She takes my hand, squeezes hard, then lets go. "Tell you what," she says. "I'm in the lab at Notre Dame. If they don't fix it in Boston, come back and let me know. I'll take a look." She grins a mischievous grin. "If worse comes to worse," she says, "I'll take it all apart and start over. Then you can stay with me, service yours truly for awhile."

I'm sipping wine, biting my lip, trying to keep from looking at her. I can't stop thinking about her being in the chair and I don't want her to see any trace of fear on my face.

"So what's with the chair?" I ask. "If you don't mind me asking, I'd like to know what happened."

"Nothing much," she says. "A transport to Moscow didn't quite take and I lost the use of my legs. It's no big deal, though, I've still got feeling down there... You'd still be able to drive me wild."

"I'd like that, Doc, but..."

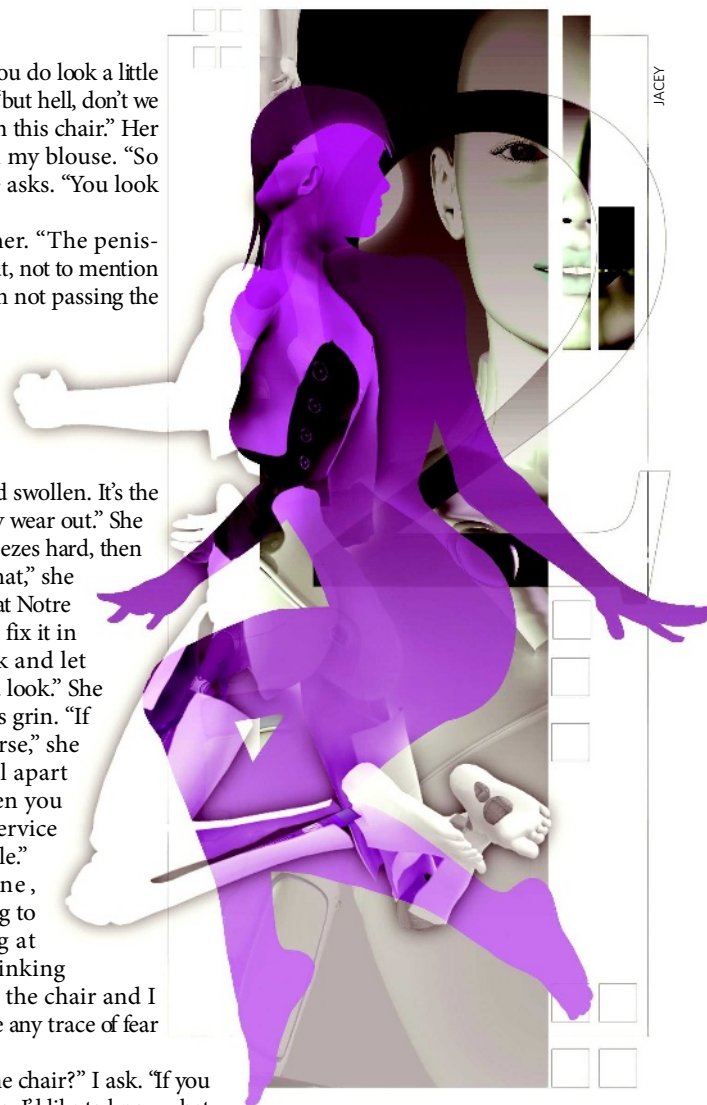
"No butts," she says. "I'll clear it with Congress, let 'em know you're going to be with me." She glances at her watch, says she's got to run. "I'll be looking for you, T-3." She smiles that cute little smile of hers, waves and keeps waving as she hovers away. I assume she's hurrying to catch a transfer. "Doc, wait," I shout, but she's already gone.

My own transport doesn't happen for another hour, so I just sit there drinking wine. I'm thinking about how Doc is in a hoverchair, and figure if she can't fix herself, how in the hell is she going to fix me. True, she's the one who turned me on, but

that was 20 years ago. It's not that I don't want to be at her beck and call, because in a way I always have been. But Doc Tan has always been well connected with the anti-science, anti-sex types in Congress. Besides, if Doc is at Notre Dame, a man's world, it's my guess she'll simply take me apart, have me fed to the hogs, piece by piece, and start over. I'm not going to let her do it, no matter what. While I'm no longer a perfect toy for men, I'm still pretty good when it comes to women. I just need fixing, that's all.

Sorry, Doc, but I just don't trust you. For one thing, the inside of what you first designed and put together is now all screwed up. For another, I'm Tan Number 2, not Tan Number 3. T-2, not T-3. ■

Janett L. Grady is a senior citizen who writes from Palmer, Alaska. Her work has appeared in a variety of magazines and newspapers. This story is her first attempt at writing science fiction.



JACEY